

PATENT COOPERATION TREATY
PCT
INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

REC'D 14 JUN 2000

WHP PCT

| | | |
|---|---|--|
| Applicant's or agent's file reference FP11010 | FOR FURTHER ACTION | See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416). |
| International application No. PCT/AU99/00385 | International filing date (<i>day/month/year</i>) 21 May 1999 | Priority Date (<i>day/month/year</i>) 21 May 1998 |
| International Patent Classification (IPC) or national classification and IPC Int. Cl. ⁷ C07H 21/04, C12Q 1/10, 1/68, C12R 1:19, G01N 37/00 | | |
| Applicant THE UNIVERSITY OF SYDNEY | | |

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of 3 sheets, including this cover sheet.
- ☒ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 6 sheet(s).

3. This report contains indications relating to the following items:
- | | | |
|------|-------------------------------------|---|
| I | <input checked="" type="checkbox"/> | Basis of the report |
| II | <input type="checkbox"/> | Priority |
| III | <input type="checkbox"/> | Non-establishment of opinion with regard to novelty, inventive step and industrial applicability |
| IV | <input type="checkbox"/> | Lack of unity of invention |
| V | <input checked="" type="checkbox"/> | Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement |
| VI | <input type="checkbox"/> | Certain documents cited |
| VII | <input type="checkbox"/> | Certain defects in the international application |
| VIII | <input type="checkbox"/> | Certain observations on the international application |

| | |
|--|--|
| Date of submission of the demand 9 December 1999 | Date of completion of the report 2 June 2000 |
| Name and mailing address of the IPEA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustalia.gov.au Facsimile No. (02) 6285 3929 | Authorized Officer MATTHEW FRANCIS Telephone No. (02) 6283 2424 |

I. Basis of the report

1. With regard to the elements of the international application:*
- ☐ the international application as originally filed.
- ☒ the description, pages 1-82, as originally filed,
pages , filed with the demand,
pages , received on with the letter of
- ☒ the claims, pages , as originally filed,
pages , as amended (together with any statement) under Article 19,
pages , filed with the demand,
pages 83-88 , received on 1 May 2000 with the letter of 27 April 2000
- ☒ the drawings, pages 1-96, as originally filed,
pages , filed with the demand,
pages , received on with the letter of
- ☒ the sequence listing part of the description:
pages 1-56, as originally filed
pages , filed with the demand
pages , received on with the letter of
2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.
These elements were available or furnished to this Authority in the following language which is:
- ☐ the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).
3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, was on the basis of the sequence listing:
- ☐ contained in the international application in written form.
- ☒ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished
4. ☐ The amendments have resulted in the cancellation of:
- ☐ the description, pages
- ☐ the claims, Nos.
- ☐ the drawings, sheets/fig.
5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

* Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).

** Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**1. Statement**

| | | |
|-------------------------------|-------------|-----|
| Novelty (N) | Claims 1-25 | YES |
| | Claims none | NO |
| Inventive step (IS) | Claims 1-25 | YES |
| | Claims none | NO |
| Industrial applicability (IA) | Claims 1-25 | YES |
| | Claims none | NO |

2. Citations and explanations (Rule 70.7)**NOVELTY (N), INVENTIVE STEP (IS):**

Claims 1-25: None of the cited art discloses or suggests nucleic acid molecules within the scope of new claims 1-6 and the claims must therefore be considered novel and inventive. Equally, claims 7-25 are both novel and inventive by virtue of their appendancy to the above claims.

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Assistant Commissioner for Patents
United States Patent and Trademark
Office
Box PCT
Washington, D.C.20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

| | |
|---|---|
| Date of mailing (day/month/year) 14 January 2000 (14.01.00) | Applicant's or agent's file reference TPG:AS:FP11010 |
| International application No. PCT/AU99/00385 | Priority date (day/month/year) 21 May 1998 (21.05.98) |
| International filing date (day/month/year) 21 May 1999 (21.05.99) | |
| Applicant REEVES, Peter, Richard et al | |

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:

09 December 1999 (09.12.99)

☐ in a notice effecting later election filed with the International Bureau on:2. The election ☒ was☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Olivia RANAIVOJAONA

Telephone No.: (41-22) 338.83.38

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 99/00385

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl⁶: C07H 21/04, (C12Q 1/10, 1/68, C12R 1:19), G01N 37/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN: Chemical Abstracts, Medline and Derwent World Patents Index using keywords e. coli and flagellin
ANGIS: Sequences corresponding to Fig 25, Fig 10 and Fig 34 (random sample)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| PX | JOURNAL OF BACTERIOLOGY, (1999 Jan) 181 (1), Reid S D; Selander R K; Whittam T S, "Sequence diversity of flagellin (fliC) alleles in pathogenic Escherichia coli", 153-160 | 1-33 |
| X | MOLECULAR MICROBIOLOGY (1994) 12(2), Tominaga A; Mahmoud T; Enomoto M, "Molecular characterization of intact, but cryptic, flagellin genes in the genus Shigella", 277-285 | 1-33 |
| X | JOURNAL OF MOLECULAR BIOLOGY (1994) 238, Fahrner K A; Block S M; Krishnaswamy S; Parkinson J S; Berg H C, "A mutant hook-associated protein (HAP3) facilitates torsionally induced transformations of the flagellar filament", 173-186 | 1-33 |

☒ Further documents are listed in the continuation of Box C

☐ See patent family annex

| | |
|---|--|
| * Special categories of cited documents: | |
| "A" document defining the general state of the art which is not considered to be of particular relevance | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "E" earlier application or patent but published on or after the international filing date | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" document referring to an oral disclosure, use, exhibition or other means | "&" document member of the same patent family |
| "P" document published prior to the international filing date but later than the priority date claimed | |

Date of the actual completion of the international search
27 July 1999

Date of mailing of the international search report
- 3 AUG 1999

Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200
WODEN ACT 2606
AUSTRALIA
Facsimile No.: (02) 6285 3929

Authorized officer

MATTHEW FRANCIS
Telephone No.: (02) 6283 2424

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|---|--|
| (51) International Patent Classification ⁶ : C07H 21/04 // (C12Q 1/10, 1:68, C12R 1:19), G01N 37/00 | A1 | (11) International Publication Number: WO 99/61458 (43) International Publication Date: 2 December 1999 (02.12.99) |
| (21) International Application Number: PCT/AU99/00385 (22) International Filing Date: 21 May 1999 (21.05.99) (30) Priority Data: PP 3634 21 May 1998 (21.05.98) AU (71) Applicant (for all designated States except US): THE UNIVERSITY OF SYDNEY [AU/AU]; Parramatta Road, Sydney, NSW 2006 (AU). (72) Inventors; and (75) Inventors/Applicants (for US only): REEVES, Peter, Richard [GB/AU]; 20 Mansfield Street, Glebe, NSW 2037 (AU). WANG, Lei [AU/AU]; 8A Holt Street, North Ryde, NSW 2113 (AU). (74) Agent: GRIFFITH HACK; G.P.O. Box 4164, Sydney, NSW 2001 (AU). | (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>1 W.O. amend claim page 83-84</i> | |
| (54) Title: ANTIGENS AND THEIR DETECTION | | |
| (57) Abstract | | |
| <p>The invention provides novel nucleotide sequences located in a gene which encodes a bacterial flagellin antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular flagellin antigens, on the basis of that antigen alone, or in conjunction with the O antigen expressed by that strain.</p> | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Larvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | Republic of Korea | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |



WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|---|--|---|--|
| (51) International Patent Classification ⁶ : C07H 21/04 // (C12Q 1/10, 1:68, C12R 1:19), G01N 37/00 | | A1 | (11) International Publication Number: WO 99/61458 (43) International Publication Date: 2 December 1999 (02.12.99) |
| (21) International Application Number: PCT/AU99/00385 (22) International Filing Date: 21 May 1999 (21.05.99) (30) Priority Data: PP 3634 21 May 1998 (21.05.98) AU (71) Applicant (for all designated States except US): THE UNIVERSITY OF SYDNEY [AU/AU]; Parramatta Road, Sydney, NSW 2006 (AU). (72) Inventors; and (75) Inventors/Applicants (for US only): REEVES, Peter, Richard [GB/AU]; 20 Mansfield Street, Glebe, NSW 2037 (AU). WANG, Lei [AU/AU]; 8A Holt Street, North Ryde, NSW 2113 (AU). (74) Agent: GRIFFITH HACK; G.P.O. Box 4164, Sydney, NSW 2001 (AU). | | (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> | |
| (54) Title: ANTIGENS AND THEIR DETECTION | | | |
| (57) Abstract <p>The invention provides novel nucleotide sequences located in a gene which encodes a bacterial flagellin antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular flagellin antigens, on the basis of that antigen alone, or in conjunction with the O antigen expressed by that strain.</p> | | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|---------------------|----|-----------------------|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | Republic of Macedonia | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's | NZ | New Zealand | | |
| CM | Cameroon | | Republic of Korea | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

Antigens and Their Detection

TECHNICAL FIELD

The invention relates to novel nucleotide sequences located in a gene which encodes a bacterial flagellin antigen, and the use of those nucleotide sequences for the detection of bacteria which express particular flagellin antigens, on the basis of that antigen alone, or in conjunction with the O antigen expressed by that strain.

BACKGROUND ART

The flagellum of many bacteria appears to be made up of a single protein known as flagellin. The serotyping schemes of *E. coli* and *Salmonella enterica* are based on highly variable antigenic surface structures which include the lipopolysaccharide which carries the O antigen and flagellin which is now known to be the carrier of the classical H antigen. In many strains of *S. enterica* there are two loci (*fliC* and *fljB*) which encode flagellin, and a regulatory system which allows one only to be expressed at any time; and which also provides for expression to rapidly alternate between the two forms first identified as two phases (H1 and H2) for the H antigen of most strains. In *E. coli* there are 54 forms of H antigen recognised and until recently they were all thought to be encoded at the *fliC* locus, as has been shown for *E. coli* K-12. However in the 1980s Ratiner [Ratiner Y A "Phase variation of the H antigen in *Escherichia coli* strain Bi327-41, the standard strain for *Escherichia coli* flagellin antigen H3" FEMS Microbiol. Lett 15 (1982) 33-36; Ratiner Y A "Presence of two structural genes determining antigenically different phase-specific flagellins in some *Escherichia coli* strains" FEMS Microbiol. Lett. 19 (1983) 37-41; Ratiner Y A "Two genetic arrangements determining flagellin antigen specificities in two diphasic *Escherichia coli* strains" FEMS Microbiol. Lett. 29 (1985) 317-323; Ratiner Y A "Different alleles of the flagellin gene *hagB* in *Escherichia coli* standard H

test strains" FEMS Microbiol Lett. 48 (1987) 97-104.] showed that in some cases there are two loci and that expression can alternate. The matter was further complicated by a recent paper by Ratiner [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] showing three loci (*flk*, *fll* and *flm*) for flagellin in addition to *fliC* although the *fljB* locus has not been found in *E. coli*. However *E. coli* strains are normally identified by the combination of one O antigen and one H antigen [and K antigen when present as a capsule (K) antigen], with no problems reported for the vast majority of cases with alternate phases, while *S. enterica* strains are normally identified by the combination of O, H1 and H2 antigens. It is still not clear how widespread in *E. coli* H antigens determined by flagellin genes other than *fliC* are.

Typing is typically carried out using specific antisera. The incidence of pathogenic *E. coli* in association with human and animal disease supports the need for suitable and rapid typing techniques.

DESCRIPTION OF THE INVENTION

In a first aspect, the present invention provides a novel nucleic acid molecule encoding all or part of an *E. coli* flagellin protein.

The present invention provides, for the first time, full length sequence for a flagellin gene for the following *E. coli* type strains: H6 (SEQ ID NO: 8), H9 (SEQ ID NO: 11), H10 (SEQ ID NO: 12), H14 (SEQ ID NO: 15), H18 (SEQ ID NO: 18), H23 (SEQ ID NO: 22), H51 (SEQ ID NO: 50), H45 (SEQ ID NO: 43), H49 (SEQ ID NO: 48), H19 (SEQ ID NO: 19), H30 (SEQ ID NO: 29), H32 (SEQ ID NO: 31), H26 (SEQ ID NO: 25), H41 (SEQ ID NO: 39), H15 (SEQ ID NO: 16), H20 (SEQ ID NO: 20), H28 (SEQ ID NO: 27), H46 (SEQ ID NO: 44), H31 (SEQ ID NO: 30), H34 (SEQ ID NO: 33), H43 (SEQ ID NO: 41) and H52 (SEQ ID NO: 51). Corrected full length sequences have been obtained for H7 (SEQ ID NO: 9) and

H12(SEQ ID NO: 14) type strains.

Partial flagellin gene sequence, including the central variable region, has been obtained for the following *E. coli* H type strains: H40(SEQ ID NO: 38), H8(SEQ ID NO: 10), H21(SEQ ID NO: 21), H47(SEQ ID NO: 46), H11(SEQ ID NO: 13), H17(SEQ ID NO: 17), H25(SEQ ID NO: 24), H42(SEQ ID NO: 40), H27(SEQ ID NO: 26), H35(SEQ ID NO: 34), H2(SEQ ID NO: 67), H3(SEQ ID NO: 68), H24(SEQ ID NO: 23), H37(SEQ ID NO: 35), H50(SEQ ID NO: 49), H4(SEQ ID NO: 6), H44(SEQ ID NO: 42), H38(SEQ ID NO: 36), H39(SEQ ID NO: 37), H55(SEQ ID NO: 53), H29(SEQ ID NO: 28), H33(SEQ ID NO: 32), H5(SEQ ID NO: 7), H54(SEQ ID NO: 52) and H56(SEQ ID NO: 54).

Comparison of sequences demonstrates that unique flagellin genes have now been sequenced (partially or completely) for the following *E. coli* H type strains: H1, H2, H3, H5, H6, H7, H9, H11, H12, H14, H15, H18, H19, H20, H21, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H37, H38, H39, H41, H42, H43, H45, H46, H48, H49, H51, H52, H54, and H56 and either H8 or H40, H10 or H50 and H4 or H17.

By comparison of these sequences, the present inventors were able to identify specific sequences for each of the above H serotypes.

The present invention also provides *fliC* sequences from 10 different H7 strains, in addition to that from the H7 type strain, and two sequences specific to H7 of O157 and O55 *E. coli* strains.

The present invention encompasses all or part of the flagellin genes sequenced for H2, H3, H5, H6, H9, H11, H14, H18, H19, H20, H21, H23, H24, H25, H26, H27, H28, H29, H30, H31, H32, H33, H34, H35, H37, H38, H39, H41, H42, H43, H44, H45, H46, H47, H48, H49, H51, H52, H54, H55, H56, H8, H40, H15, H10, or H50, H4 and H17 type strains. Of these flagellin genes sequenced, those from the type strains for H8 and H40 are identical, those from type strains H10 and H50, H1 and H12, H38 and H55, H21 and

H47, and H4, H17 and H44 type strains are highly similar.

The invention also encompasses newly provided sequence for H7 and H12 as well as novel primers for the specific amplification of H1, H7, H12 and H48 as well as for the other above mentioned newly sequenced flagellin genes.

By cloning and expression of these sequenced flagellin genes in a *fliC* deletion *E. coli* K-12 strain, and use of anti-H antiserum, we have confirmed the H specificities encoded by 39 flagellin genes. The 39 H specificities are H1, H2, H4, H5, H6, H7, H9, H10, H11, H12, H14, H15, H16, H18, H19, H20, H21, H23, H24, H26, H27, H28, H29, H30, H31, H32, H33, H34, H38, H39, H41, H42, H43, H45, H46, H49, H51, H52, and H56, encoded by flagellin genes obtained from H type strains for H1, H2, H4, H5, H6, H7, H9, H10, H11, H12, H14, H15, H3, H18, H19, H20, H21, H23, H24, H26, H27, H28, H29, H30, H31, H32, H33, H34, H38, H39, H41, H42, H43, H45, H46, H49, H51, H52, and H56 respectively.

The nucleic acid molecules of the invention may be variable in length. In one embodiment they are oligonucleotides of from about 10 to about 20 nucleotides in length. The oligonucleotides of the invention are specific for the flagellin gene from which they are derived and are derived from the central region of the gene. In one embodiment, oligonucleotides in accordance with the present invention, which also include oligonucleotides from the previously sequenced *E. coli* H1, H7, H12 and H48 genes, are those shown in Table 3.

The 45 sequences (see Table 3) provide a panel to which newly sequenced genes can be compared to select specific oligonucleotides for those newly sequenced genes.

In a second aspect the invention provides a method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one nucleic acid molecule derived from a flagellin gene, wherein the at

least one nucleic acid molecule is specific for a particular flagellin gene associated with the H serotype, to any *E. coli* in the sample which contain the gene, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

In one preferred embodiment the detection method is a Southern blot method. More preferably, the nucleic acid molecule is labelled and hybridisation of the nucleic acid molecule is detected by autoradiography or detection of fluorescence.

Preferred nucleic acid molecules for the detection of particular flagellin genes are listed in Table 3.

In a third aspect the invention provides a method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one pair of nucleic acid molecules to any *E. coli* in the sample which contains the flagellin gene for the particular H serotype, wherein at least one of the nucleic acid molecules is specific for the particular flagellin gene associated with the H serotype, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

In one preferred embodiment the detection method is a polymerase chain reaction method. More preferably, the nucleic acid molecules are labelled and hybridisation of the nucleic acid molecule is detected by electrophoresis.

It is recognised that there may be instances where spurious hybridisation will arise through the initial selection of a sequence found in many different genes but this is typically recognisable by, for instance, comparison of band sizes against controls in PCR gels, and an alternative sequence can be selected.

In a fourth aspect the invention provides a method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

5 (a) specifically hybridising at least one nucleic acid molecule, derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the
10 synthesis of a particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

(b) specifically hybridising at least one nucleic acid molecule derived from and specific for a particular flagellin gene associated with that H serotype, to any *E.*
15 *coli* in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic acid molecules.

Preferred nucleic acid molecules for the detection of particular flagellin genes are listed in Table 3.

20 In one preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O111 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH*
25 (nucleotide position 739 to 1932 of Figure 5), *wzx* (nucleotide position 8646 to 9911 of Figure 5), *wzy* (nucleotide position 9901 to 10953 of Figure 5), *wbdM* (nucleotide position 11821 to 12945 of Figure 5) and fragments of those molecules of at least 10-12 nucleotides
30 in length. Particularly preferred nucleic acid molecules are those set out in Tables 8 and 8A, with respect to the above mentioned genes.

In another preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is
35 specific to the nucleotide sequence encoding the O157 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdN*

(nucleotide position 79 to 861 of Figure 6), *wbdO*
(nucleotide position 2011 to 2757 of Figure 6), *wbdP*
(nucleotide position 5257 to 6471 of Figure 6), *wbdR*
(nucleotide position 13156 to 13821 of Figure 6), *wzx*
5 (nucleotide position 2744 to 4135 of Figure 6) and *wzy*
(nucleotide position 858 to 2042 of Figure 6) and
fragments of those molecules of at least 10-12 nucleotides
in length. Particularly preferred nucleic acid molecules
are those set out in Tables 9 and 9A, with respect to the
10 above mentioned genes.

In one preferred embodiment the detection method is a
Southern blot method. More preferably, the nucleic acid
molecule is labelled and hybridisation of the nucleic acid
molecule is detected by autoradiography or detection of
15 fluorescence.

In a fifth aspect the invention provides a method for
detecting the presence of a particular O serotype and H
serotype of *E. coli* in a sample, the method comprising the
following steps:

20 (a) specifically hybridising at least one pair of
nucleic acid molecules, at least one of which is derived
from and specific for a gene encoding a transferase or a
gene encoding an enzyme for the transport or processing of
a polysaccharide or oligosaccharide unit, the gene being
25 involved in the synthesis of the particular *E. coli* O
antigen, to any *E. coli* in the sample which contain the
gene;

(b) specifically hybridising at least one pair of
nucleic acid molecules, at least one of which is derived
30 from and specific for a particular flagellin gene
associated with the particular H serotype, to any *E. coli*
in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic
acid molecules.

35 Preferred nucleic acid molecules for the detection of
particular flagellin genes are listed in Table 3.

In one preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O111 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdH* (nucleotide position 739 to 1932 of Figure 5), *wzx* (nucleotide position 8646 to 9911 of Figure 5), *wzy* (nucleotide position 9901 to 10953 of Figure 5), *wbdM* (nucleotide position 11821 to 12945 of Figure 5) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 8 and 8A, with respect to the above mentioned genes.

In another preferred embodiment, the sequence of the nucleic acid molecule specific for the O antigen is specific to the nucleotide sequence encoding the O157 antigen. More preferably, the sequence is derived from a gene selected from the group consisting of *wbdN* (nucleotide position 79 to 861 of Figure 6), *wbdO* (nucleotide position 2011 to 2757 of Figure 6), *wbdP* (nucleotide position 5257 to 6471 of Figure 6), *wbdR* (nucleotide position 13156 to 13821 of Figure 6), *wzx* (nucleotide position 2744 to 4135 of Figure 6) and *wzy* (nucleotide position 858 to 2042 of Figure 6) and fragments of those molecules of at least 10-12 nucleotides in length. Particularly preferred nucleic acid molecules are those set out in Tables 9 and 9A, with respect to the above mentioned genes.

In one preferred embodiment the detection method is a polymerase chain reaction method. More preferably, the nucleic acid molecules are labelled and hybridisation of the nucleic acid molecule is detected by electrophoresis.

The present inventors believe that based on the teachings of the present invention and available information concerning O antigen gene clusters, and through use of experimental analysis, comparison of nucleic acid sequences or predicted protein structures, nucleic acid molecules in accordance with the invention

can be readily derived for any particular O antigen of interest. Suitable bacterial strains can typically be acquired commercially from depositary institutions.

There are currently 166 defined *E. coli* O antigens.

5 Samples of the 166 different *E. coli* O antigen serotypes are available from Statens Serum Institut, Copenhagen, Denmark.

10 The inventors envisage rare circumstances whereby two genetically similar gene clusters encoding serologically different O antigens have arisen through recombination of genes or mutation so as to generate polymorphic variants.

15 In these circumstances multiple pairs of oligonucleotides may be selected to provide hybridisation to the specific combination of genes. The invention thus envisages the use of a panel containing multiple nucleic acid molecules for use in the method of testing for O antigen in conjunction with H antigen, wherein the nucleic acid molecules are derived from genes encoding transferases and/or enzymes for the transport or processing of a polysaccharide or oligosaccharide unit including *wzx* or *wzy* genes, wherein the panel of nucleic acid molecules is specific to a particular O antigen. The panel of nucleic acid molecules can include nucleic acid molecules derived from O antigen sugar pathway genes where necessary.

25 The inventors also found two mutated flagellin genes from H type strains for H35 and H54 which have insertion sequences inserted into normal flagellar genes identical or near identical to that of the H11 and H21 type strains respectively. Thus, primers for H11 and H21 (listed in Table 3) would also amplify fragments in H35 and H54, which differ in sizes to those in H11 and H21 respectively. The inventors also provide two pairs of primers each for H35 and H54 based on the insertion sequence (see H35 and H54 columns in Table 3). The use of one of them in combination with one of the H11 or H21 primers will generate a PCR band only in H35 or H54 respectively, and this will also differentiate H35 and H54

from H11 and H21 respectively.

The present invention also relates to methods of detecting the presence of particular *E. coli* H antigens or H antigen and O antigen combinations where one or more nucleic acid molecules which generate a particular size fragment indicative of the presence of that H antigen are used or in which the combination of one antigen specific primer for that H antigen with another primer for a related H antigen provides for the detection of the particular H antigen by hybridisation to the relevant gene. Preferably, the H antigen is H11, H21, H35 or H54.

The pairs of nucleic acid molecules where the method of the fifth aspect is used may both hybridise to the relevant H or O antigen gene or alternatively only one may hybridise to the relevant gene and the other to another site.

The inventors recognise in applying the methods of the invention for detecting combinations of O and H antigens to samples, that the methods do not indicate whether a positive result for a particular O and H antigen combination arises because the O and H antigen are present on a single *E. coli* strain present in the sample or are present on different *E. coli* strains present in the sample. Because the ability to identify the presence of *E. coli* strains with particular O and H antigen combinations is highly desirable (due to the relationship between particular combinations and pathogenicity) the determination that a particular combination is present in a sample can be followed by isolation of single colonies and checking whether they contain the relevant combination by using the same method again or using antibody labelled magnetic beads to separate cells expressing the particular O or H antigen and then testing the isolated cells for the other serotype.

In addition, as mentioned above, the present inventors have established the existence of H7 primers specific to the O157 and O55 serotypes. Using such

primers it is possible to detect particular O and H antigen combinations with the use of H specific nucleic acid molecules.

In a sixth aspect the invention provides a method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

(a) specifically hybridising at least one nucleic acid molecule, derived from and specific for a gene encoding a flagellin associated with a particular *E. coli* H antigen serotype to any *E. coli* carrying the gene and present in the sample;
and

(b) detecting the at least one specifically hybridised nucleic acid molecule, wherein the at least one nucleic acid molecule is specific for the particular combination of O and H antigen.

Preferably the combination is O55:H7 or O157:H7.

The ability to detect the O157:H7 combination from a particular H7 primer or pair is of particular use given the association of this combination with pathogenic strains.

In a seventh aspect the present invention provides a method for testing a food derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

In an eighth aspect the present invention provides a method for testing a faecal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

In a ninth aspect the present invention provides a method for testing a patient or animal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens comprising testing the sample by a method of the fourth, fifth or sixth aspect the invention.

Preferably, the method of the seventh, eighth or ninth aspect of the invention is a polymerase chain reaction method. More preferably the oligonucleotide molecules for use in the method are labelled. Even more preferably the hybridised nucleic acid molecules are detected by electrophoresis.

In the above described methods it will be understood that where pairs of nucleic acid molecules are used one of the nucleic acid molecules may hybridise to a sequence that is not from the O antigen transferase, *wzx* or *wzy* gene or the flagellin gene. Further where both hybridise to these genes the O antigen molecules may hybridise to the same or a different one of these genes.

In a tenth aspect the present invention provides a kit for identifying the H serotype of *E. coli*, the kit comprising:

at least one nucleic acid molecule derived from and specific for an *E. coli* flagellin gene.

In an eleventh aspect the present invention provides a kit for identifying the H and O serotype of *E. coli*, the kit comprising:

(a) at least one nucleic acid molecule derived from and specific for an *E. coli* flagellin gene; and

(b) at least one nucleic acid molecule derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen.

The nucleic acid molecules may be provided in the same or different vials. The kit may also provide in the same or separate vials a second set of specific nucleic acid molecules.

Particularly preferred nucleic acid molecules for inclusion in the kits are those specified in Tables 3, 8, 8A, 9 and 9A as described above.

DEFINITIONS

In this specification, we have used term "flagellin gene" in many cases where previously one would have used "*fliC*", to allow for the uncertainty as to locus introduced by recent observations. However, uncertainty as to the locus does not alter the fact that most *E. coli* strains express a single H antigen and that a single flagellin gene sequence per strain is required to give the genetic basis for H antigen variation. Any use of the name *fliC* in this specification where a different locus is later shown to be involved would not affect the validity of conclusions drawn regarding application of information based on the sequence, where the conclusions do not relate to the map position. Thus it is generally the nucleic acid molecule itself which is of importance rather than the name attributed to the gene. When it is known or suspected that the gene encoding the H antigen is not in the *fliC* locus, we use the term flagellin rather than *fliC*.

The phrase, "a nucleic acid molecule derived from a gene" means that the nucleic acid molecule has a nucleotide sequence which is either identical or substantially similar to all or part of the identified gene. Thus a nucleic acid molecule derived from a gene can be a molecule which is isolated from the identified gene by physical separation from that gene, or a molecule which is artificially synthesised and has a nucleotide sequence which is either identical to or substantially similar to all or part of the identified gene. While some workers consider only the DNA strand with the same sequence as the mRNA transcribed from the gene, here either strand is intended.

Transferase genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that transfer monomeric sugar units.

Flippase or *wzx* genes are regions of nucleic acid which have a nucleotide sequence which encodes a gene

product that flips oligosaccharide repeat units generally composed of three to six monomeric sugar units to the external surface of the membrane.

5 Polymerase or *wzy* genes are regions of nucleic acid which have a nucleotide sequence which encodes gene products that polymerise repeating oligosaccharide units generally composed of 3-6 monomeric sugar units.

10 The nucleotide sequences provided in this specification are described as anti-sense sequences. This term is used in the same manner as it is used in Glossary of Biochemistry and Molecular Biology Revised Edition, David M. Glick, 1997 Portland Press Ltd., London on page 11 where the term is described as referring to one of the two strands of double-stranded DNA usually that which has
15 the same sequence as the mRNA. We use it to describe this strand which has the same sequence as the mRNA.

NOMENCLATURE

Synonyms for *E. coli* O111 *rfb*

| | <u>Current names</u> | <u>Our names</u> | <u>Bastin et al. 1991</u> |
|----|----------------------|------------------|---------------------------|
| 20 | wbdH | orf1 | |
| | gmd | orf2 | |
| | wbdI | orf3 | orf3.4* |
| | manC | orf4 | rfbM* |
| | manB | orf5 | rfbK* |
| 25 | wbdJ | orf6 | orf6.7* |
| | wbdK | orf7 | orf7.7* |
| | wzx | orf8 | orf8.9 and rfbX* |
| | wzy | orf9 | |
| | wbdL | orf10 | |
| 30 | wbdM | orf11 | |

* Nomenclature according to Bastin D.A., et al. 1991 "Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231.

Other Synonyms

| | | |
|----|------|-------------------------------------|
| | wzy | rfc |
| | wzx | rfbX |
| | rmlA | rfbA |
| 40 | rmlB | rfbB |
| | rmlC | rfbC |
| | rmlD | rfbD |
| | glf | orf6* |
| | wbbI | orf3#, orf8* of <i>E. coli</i> K-12 |

wbbJ orf2#, orf9* of E. coli K-12
wbbK orf1#, orf10* of E. coli K-12
wbbL orf5#, orf 11* of E. coli K-12
Nomenclature according to Yao, Z. And M. A. Valvano 1994.

5 "Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes the confer groups-specificity to Shigella flexneri serotypes Y and 4a". J. Bacteriol. 176: 4133-4143.

10 * Nomenclature according to Stevenson et al. 1994. "Structure of the O-antigen of E. coli K-12 and the sequence of its rfb gene cluster". J. Bacteriol 176: 4144-4156.

15 • The O antigen genes of many species were given rfb names (rfbA etc) and the O antigen gene cluster was often referred to as the rfb cluster. There are now new names for the rfb genes as shown in the table. Both terminologies have been used herein, depending on the source of the information.

20 In the claims that follow and in the summary of the invention, except where the context requires otherwise due to express language or necessary implication, the word "comprising" is used in the sense of "including", i.e. the features specified may be associated with further features in various embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Figure 1 shows Eco R1 restriction maps of cosmid clones pPR1054, pPR1055, pPR1056, pPR1058, pPR1287 which are subclones of E. coli O111 O antigen gene cluster. The thickened line is the region common to all clones. Broken lines show segments that are non-contiguous on the chromosome. The deduced restriction map for E. coli strain M92 is shown above.

30 Figure 2 shows a restriction mapping analysis of E. coli O111 O antigen gene cluster within the cosmid clone pPR1058. Restriction enzymes are: (B: BamHI; Bg: BglII, E: EcoRI; H: HindIII; K: KpnI; P: PstI; S: SalI and X: XhoI. Plasmids pPR1230, pPR1231, and pPR1288 are deletion derivatives of pPR1058. Plasmids pPR 1237, 35 pPR1238, pPR1239 and pPR1240 are in pUC19. Plasmids pPR1243, pPR1244, pPR1245, pPR1246 and pPR1248 are in pUC18, and pPR1292 is in pUC19. Plasmid pPR1270 is in 40

pT7T319U. Probes 1, 2 and 3 were isolated as internal fragments of pPR1246, pPR1243 and pPR1237 respectively. Dotted lines indicate that subclone DNA extends to the left of the map into attached vector.

5 Figure 3 shows the structure of *E. coli* O111 O antigen gene cluster.

 Figure 4 shows the structure of *E. coli* O157 O antigen gene cluster.

10 Figure 5 shows the nucleotide sequence (SEQ ID NO: 45) of the *E. coli* O111 O antigen gene cluster. Note: (1) The first and last three bases of a gene are underlined and of *italic* respectively.; (2) The region which was previously sequenced by Bastin and Reeves 1995 "Sequence and analysis of the O antigen gene (*rfb*) cluster of
15 *Escherichia coli* O111" Gene 164: 17-23 is marked.

 Figure 6 shows the nucleotide sequence (SEQ ID NO: 56) of the *E. coli* O157 O antigen gene cluster. Note: (1) The first and last three bases of a gene (region) are underlined and of *italic* respectively (2) The region
20 previously sequenced by Bilge et al. 1996 "Role of the *Escherichia coli* O157-H7 O side chain in adherence and analysis of an *rfb* locus". Inf. and Immun 64:4795-4801 is marked.

25 Figures 7 to 9 show the nucleotide sequences (SEQ ID NOS: 66 to 68 respectively) obtained for flagellin genes from *E. coli* type strains for H1 to H3 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

30 Figures 10 to 18 show the nucleotide sequences (SEQ ID NOS: 6 to 14 respectively) obtained for flagellin genes from *E. coli* type strains for H4 to H12 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

35 Figures 19 and 20 show the nucleotide sequences (SEQ ID NOS: 15 to 16 respectively) obtained for flagellin genes from *E. coli* type strains for H14 and H15 respectively. The primer positions listed in Table 3 are

based on treating the first nucleotide of each of these sequences as No. 1.

Figures 22 and 26 show the nucleotide sequences (SEQ ID NOS: 17 to 21 respectively) obtained for flagellin genes from *E. coli* type strains for H17 and H21 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 27 to 39 show the nucleotide sequences (SEQ ID NOS: 22 to 34) obtained for flagellin genes from *E. coli* type strains for H23 to H35 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 40 to 49 show the nucleotide sequences (SEQ ID NOS: 35 to 44) obtained for flagellin genes from *E. coli* type strains for H37 to H46 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 50 to 55 show the nucleotide sequences (SEQ ID NOS: 46 to 51) obtained for flagellin genes from *E. coli* type strains for H47 to H52 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 56 to 58 show the nucleotide sequences (SEQ ID NOS: 52 to 54) obtained for flagellin genes from *E. coli* type strains for H54 to H56 respectively. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figure 59 shows the nucleotide sequence (SEQ ID NO: 55) obtained for the flagellin gene from *E. coli* H7 strain M1179. The primer positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figures 60 to 68 show the nucleotide sequences (SEQ ID NOS: 57 to 65 respectively) obtained for flagellin genes from *E. coli* strains M1004, M1211, M1200, M1686, M1328, M917, M527, M973, and M918 respectively. The primer

positions listed in Table 3 are based on treating the first nucleotide of each of these sequences as No. 1.

Figure 69 shows the nucleotide sequence (SEQ ID NO: 1) of the *fliC* gene and DNA flanking the *fliC* gene from the H25 type strain.

Figure 70A shows the nucleotide sequence (SEQ ID NO: 2) obtained from the 5' end of the insert of plasmid pPR1989. The insert of plasmid pPR1989 encodes the second flagellin gene of the H55 type strain.

Figure 70B shows the nucleotide sequence (SEQ ID NO: 3) obtained from the 3' end of the insert of plasmid pPR1989. The insert of plasmid pPR1989 encodes the second flagellin gene of the H55 type strain.

Figure 71 shows the nucleotide sequence (SEQ ID NO:4) obtained from the 5' end of the insert of plasmid pPR1993.

The insert of plasmid pPR1993 encodes the second flagellin gene of the H36 strain.

Figure 72 shows the nucleotide sequence (SEQ ID NO:5) obtained from the 3' end of the insert of plasmid pPR1993.

The insert of plasmid pPR1993 encodes the second flagellin gene of the H36 type strain.

Figure 73 A shows the sequence of polylinker and the SD sequence of plasmid pTrc99A.

Figure 73B shows the sequence of the junction region between the SD sequence and the start of flagellin gene in the plasmids used for the expression of flagellin genes.

BEST METHOD OF CARRYING OUT THE INVENTION

In carrying out the methods of the invention with respect to the testing of particular sample types including samples from food, patients, animals and faeces the samples are prepared by routine techniques routinely used in the preparation of such samples for DNA based testing. The steps for testing the samples using particular nucleic acid molecules in assay formats such as Southern blots and PCR are performed under routinely determined conditions appropriate to the sample and the

nucleic acid molecules.

H antigen

Materials and Methods

5 1. *Bacterial strains and plasmid:*

There are 54 H types in *E. coli* [Ewing, W.H.: Edwards and Ewing's identification of the *Enterobacteriaceae*., Elsevier Science Publishers, Amsterdam, The Netherlands, 1986]: note H antigens from 1 to 57 were listed and that
10 13, 22 and 57 are not valid. All the standard H type strains except H16 were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia. The primary stocks are hold at the Statens Serum Institut, Copenhagen, Denmark.

15 The additional H7 strains used are listed in Table 1.

We do not have the type strain for H16. It is known that the H3 type strain is biphasic and can also express the H16 flagellin gene [Ratiner, Y. A. (1985) "Two genetic arrangements determining flagellar antigen specificities
20 in two diphasic *E. coli* strains. FEMS Microbiol Lett 19: 317-323]. We have sequenced and cloned the H16 flagellin gene from the H3 type strain (see below).

E. coli K-12 strain C600 *hsm hsr fliC::Tn10* [Kuwayiwa, G. (1988) "Flagellin domain that affects H
25 antigenicity of *E. coli* K-12" J. Bacteriol. 170; 485-488] (laboratory stock no. M2126) was obtained from Dr Benita Westerlund-Wikstrom of the Department of Biosciences, University of Helsinki, Finland. *E. coli* K-12 strain EJ2282 (laboratory no. P5560) is a *fliC* deletion strain,
30 and was obtained from Dr Masatoshi Enomoto of the Department of Biology, Okayama University, Japan [Tominaga, A. M. A.-H. Mahmoud, T. Mokaiharu and M. Enomoto (1994) "Molecular characterization of intact but cryptic, flagellin genes in the genus *Shigella*.: Mol.
35 Microbiol. 12: 277-285].

Plasmid pTrc99A was purchased from Pharmacia LKB (Melbourne, VIC, Australia).

2. Antisera

Antisera against H1, H3, H8, H14, H15, H17, H23, H24, H25, H26, H29, H30, H31, H32, H33, H35, H36, H37, H38, H39, H43, H44, H46, H47, H48, H49, H52, H53, H54, H55, and H56 were obtained from the Institute of Medical and veterinary Science, Adelaide, Australia. Antisera against H2, H4, H5, H6, H7, H9, H10, H11, H12, H16, H18, H19, H20, H21, H27, H28, H34, H40, H41, H42, H45, and H51 were obtained from Denka Seiken Co., Ltd, Tokyo, Japan.

Antisera to type H50 was not available from any known source.

The antisera available were checked against the appropriate type strains to confirm the specificities of both flagellin H antigen and H antisera: 52 sera (all those except anti-H16 serum listed above) gave a positive reaction with the corresponding type strains for that serum.

3. Agglutination test:

Bacteria from 1 ml of an overnight culture grown in Luria broth (Difco Tryptone, 10g/l; Difco yeast extract, 5g/l; NaCl, 0.5 g/l; pH 7.2) at 30°C was centrifuged (4000 rpm/10 min) and the bacteria pellet resuspended in 100 ml of saline. The agglutination test was carried out by mixing equal volumes (5 ml) of both the cells and antiserum on a slide. The slide was rocked for 1 minute and then observed for agglutination. For all agglutination tests, saline containing no antiserum was mixed with cells to be used as a negative control.

For testing the H specificities of strain M2126 or strain P5560 carrying plasmid containing cloned flagellin genes, cells of M2126 or P5560 were used as an additional negative control.

All agglutination tests were first carried out using undiluted antisera (note that the antisera we used have been diluted before reaching our hands), except for anti-

H11, anti-H34, anti-H52 and anti-H26 serum for which we used 1:10 dilutions to avoid background agglutination. In cases for which cross-reactions have been reported, we carried out agglutination tests using serial dilutions of sera (see section 10.1)

4. Motility test:

The motility of strain M2126 or strain P5560 carrying cloned flagellin genes was examined microscopically. 1 ml of overnight culture grown in Luria broth (Difco Tryptone, 10g/l; Difco yeast extract, 5g/l; NaCl, 0.5 g/l; pH 7.2) at 30°C was inoculated into 10 ml of Luria broth, and the culture was shaken at 100 rpm at 30°C to early log phase (OD 625 = 0.2). A loopful of culture was placed on a slide and examined under a microscope. Motility of individual cells was easily distinguished from Brownian movement and streaming, and presence or absence of motility recorded.

5. Isolation of chromosomal DNA:

Chromosomal DNA from all the 53 H type strains and the strains listed in Table 1 was isolated using the Promega Genomic isolation kit (Madison WI USA). Each chromosomal DNA sample was checked by gel electrophoresis of the DNA and by PCR amplification of the *mdh* gene using oligonucleotides based on the *E. coli* K-12 *mdh* gene [Boyd, E.F., Nelson, K., Wang, F.-S., Whittam, T.S. and Selander, R.K.: Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA 91 (1994) 1280-1284].

6. PCR amplification of flagellin gene:

Flagellin genes from different strains were first PCR amplified using one of the following four pairs of oligonucleotides:

#1285 (5'-atggcacaagtcattaatac) and

#1286 (5'-ttaaccctgcagtagagaca);

- 22 -

#1417 (5'-ctgatcactcaaaataatatcaac) and
#1418 (5'-ctgcggtacctgggttggc);
#1431 (5'-atggcacaagtcattaatacccaac) and
#1432 (5'-ctaaccctgcagcagagaca):
5 #1575 (5'-gggtggaaacccaatacg) and
#1576 (5'-gcgcatcaggcaatttgg)

10 PCR reactions were carried out under the following
conditions: denaturing, 94°C/30'; annealing, temperature
varies (refer to Table 2)/30'; extension, 72°C/1'; 30
cycles. The PCR product was purified using the Promega
Wizard PCR purification kit (Madison WI USA) before being
sequenced.

15 The H36 and H53 type strains gave two PCR bands using
primer pairs #1431/#1432 and #1417/#1418 respectively, and
were not sequenced.

7. Enzymes and buffers:

20 Restriction endonucleases and DNA T4 ligase were
purchased from Boehringer Mannheim (Castle Hill, NSW,
Australia). Restriction enzymes were used in the
recommended commercial buffer.

8. Sequencing of the flagellin genes:

25 Each PCR product was first sequenced using the
oligonucleotide primers used for the PCR amplification.
Primers based on the obtained sequence were then used to
sequence further, and this procedure was repeated until
the entire PCR product was sequenced.

30 The sequencing reactions were performed using the
DyeDeoxy Terminator Cycle Sequencing method (Applied
Biosystems, CA, USA), and reaction products were analysed
using fluorescent dye and an ABI377 automated sequencer
(CA, USA).

35 Sequence data were processed and analysed using
Staden programs [Sacchi CT, Zanella R C, Caugant D A,
Frasch C E, Hidalgo N T, Milagres L G, Pessoa L L, Ramos S
R, Camargo M C C and Melles C E A "Emergence of a new

clone of serogroup C *Neisseria meningitidis* in Sao Paulo, Brazil" J. Clin. Microbiol. 30 (1992) 1282-1286;

Staden, R.: Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids Res. 10 (1982a) 4731-4751;

Staden, R.: An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucl. Acids Res. 10 (1982b) 2951-2961;

Staden, R.: Computer methods to locate signals in nucleic acid sequences. Nucl. Acids Res. 12 (1984a) 505-519;

Staden, R.: Graphic methods to determine the function of nucleic acid sequences. A summary of ANALYSEQ options. Nucl. Acids Res. 12 (1984b) 521-538;

Staden, R.: The current status and portability of our sequence handling software. Nucl. Acids Res. 14 (1986) 217-231].

We were able to PCR amplify flagellin genes from H type strains for H7, 23, 12, 51, 45, 49, 19, 9, 30, 32, 26, 41, 15, 20, 28, 46, 31, 14, 18, 6, 34, 48, 43, 10, 52, and also from H7 strains m1004, m527, m1686, m1211, m1328, m973, m1179, m1200, m917, and m918 using primers #1575 and #1576 which are based on sequences 51-34 bp upstream and 37-54 bp downstream of start and end of the *E. coli* K-12 *fliC* gene respectively. Thus, the full sequence of the flagellin gene from these strains was obtained and the use of flanking sequence for primers makes it highly likely that they are at the *fliC* locus.

For other strains, we were only able to amplify the flagellin gene using one or more of the other three pairs of primers, which are based on sequence within the *fliC* gene, and thus only partial sequence was obtained. These amplicons may be of the *fliC* gene or one of the alternative flagellin genes. The flagellin gene sequences from H type strains for H40, 8, 21, 47, 11, 27, 35, 2, 3, 24, 37, 50, 4, 44, 38, 55, 29, 33, 5, and 56 obtained are lacking 18 and 14 codons at 5' and 3' ends respectively. The flagellin gene sequence of H39 obtained using primers

#1285/#1286 lacks 18 and 19 codons at 5' and 3' ends respectively. The flagellin gene sequence of H type strains of H17, 25 and 42 lack 23 and 21 codons at 5' and 3' ends respectively. The flagellin gene sequence of the H type strain for H54 lacks 23 and 12 codons at the 5' and 3' ends respectively. There is very little variation in the sequence at the two ends of flagellin genes and antigenic variation is due to variation in the central region of the gene. The absence of sequence for the ends of some of the flagellin genes is not important for the purpose of the present invention relating to the detection of antigenic variation by DNA sequence based means.

The *fliC* genes from H type strains of H1, H7 and H12 have been sequenced previously [Schoenhals, G. and Whitfield, C.: Comparative analysis of flagellin sequences from *Escherichia coli* strains possessing serologically distinct flagellar filaments with a shared complex surface pattern. J. Bacteriol. 175 (1993) 5395-5402] and we did not sequence the gene from the H1 strain.

We have sequenced *fliC* genes from a set of H7 strains with different O antigens, including that of *fliC* from the H7 type strain as one of the set: we have found four differences from the published H7 sequence (GenBank accession number L07388) which we believe are due to errors in the published sequence.

We have also re-sequenced the *fliC* gene from the H12 type strain, and have found one difference from the published H12 sequence (GenBank accession number L07389) which we believe is due to an error in the published sequence.

The flagellin genes from type strains H35 and H54 were also amplified using primers #1431/#1432, which are based on sequence within the *fliC* gene. Sequence data revealed that these two genes would be non-functional due to insertion sequence inserted in the middle of them. We have sequenced them to facilitate selection of primers for the functional flagellin genes.

9. Cloning of flagellin genes

DNA was digested for 2 hr at 37°C with appropriate restriction enzyme(s). The reaction product was then extracted once with phenol, and twice with ether. DNA was precipitated with 2 vols of ethanol and resuspended in water before the ligation reaction was carried out. Ligation was carried out O/N at 4°C and the ligated DNA was electroporated into one of the *E. coli fliC* mutant strains.

9.1. Cloning of flagellin genes from type strains for H1, H2, H3, H5, H6, H7, H9, H10, H11, H12, H14, H15, H18, H19, H20, H21, H24, H26, H27, H28, H29, H31, H34, H38, H39, H41, H42, H43, H45, H46, H49, H51, H52, and H56:

The full flagellin gene was PCR amplified using primers #1868 and #1870 (Table 3A). Both these primers are based on the sequences of the H7 flagellin gene of the H7 type strain. #1868 is the 5' primer: there is an *NcoI* site incorporated into the primer (Table 3B) and the flagellin gene starts at base 3 of the *NcoI* site. The 3' primer #1870 has a *BamHI* site incorporated downstream of the stop codon of the flagellin gene (Table 3B). PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, temperature varies (refer to Table 3A)/30'; extension, 72°C/1'; 30 cycles. The PCR product was purified using the Promega Wizard PCR purification kit (Madison WI USA) before being digested by restriction enzymes *NcoI* and *BamHI* and cloned into the *NcoI/BamHI* sites of plasmid pTrc99A.

Plasmid pTrc99A has a strong *trc* promoter upstream of the polylinker. Downstream of the promoter, it contains the ribosome binding site (SD sequence, see Fig 73) which is located 8bp upstream of the ATG site within the *NcoI* site. The polylinker and the SD sequence of pTrc99A are shown in Fig 73.

The plasmids generated were given pPR numbers, and

they are listed in Table 3A. In these plasmids, the expression module consists of the *trc* promoter, the SD sequence, and the full flagellin gene. The sequence of the junction region between the SD sequence and the start of flagellin gene is shown in Fig 73.

For flagellin genes from type strains for H6, H7, H9, H10, H12, H14, H18, H19, H20, H26, H28, H31, H41, H43, H45, H46, H49, H51, and H52, we have the full sequence for each gene and the primer sequences (#1868 and #1870) are conserved among them. The cloned genes therefore have the same sequence as those from the type strains.

For flagellin genes from type strains for H1, H15 and H34, we also have the full sequence. The previously published sequence of the flagellin gene from the H1 type strain was extracted from GenBank (accession number L07387) and used. Primer #1868 is conserved in all three. But, primer #1870 has the third base of the fifth last codon in the H1 sequence changed from A to G, and the third base of the second last codon changed from C to T in the H15 and H34 sequences: these changes did not change the amino acid coded, so the cloned genes encode the same gene products as those of the type strains.

For flagellin genes from type strains for H2, H3, H5, H11, H21, H24, H27, H29, H38, H39, H42, and H56, we do not have the full sequences. In the plasmids carrying genes from these type strains, the expression module consists of the *trc* promoter, the SD sequence, and the full flagellin gene with the first and the last 21 base pairs being determined by the primer sequences which are based on the H7 flagellin gene of the H7 type strain. The sequence of the junction region between the SD sequence and the start of flagellin gene is shown in Fig 73.

9.2. Cloning of the flagellin gene from type strain of H23:

The full flagellin gene was PCR amplified using primers #1868 and #1869 (Table 3A). #1868 is the 5'

primer: there is an *Nco*I site incorporated into the primer (Table 3B) and the flagellin gene starts at base 3 of the *Nco*I site. The 3' primer #1869 has a *Sal*I site incorporated downstream of the stop codon of the flagellin gene (Table 3B). PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, 55°C/30'; extension, 72°C/1'; 30 cycles. The PCR product was purified using the Promega Wizard PCR purification kit (Madison WI USA) before being digested by restriction enzymes *Nco*I and *Sal*I and cloned into the *Nco*I/*Sal*I sites of plasmid pTrc99A to give plasmid pPR1942.

Plasmid pTrc99A has a strong *trc* promoter upstream of the polylinker. Downstream of the promoter, it contains the ribosome binding site (SD sequence, see Fig 73) which is located 8bp upstream of the ATG site within the *Nco*I site. The polylinker and the SD sequence of pTrc99A are shown in Fig 73.

The expression module of pPR1942 consists of the *trc* promoter, the SD sequence, and the full flagellin gene. The sequence of the junction region between the SD sequence and the start of flagellin gene is shown in Fig 73.

9.3. Cloning of flagellin genes from type strains of H30, H32 and H33:

The full flagellin gene was PCR amplified using primers #1868 and #1871 (Table 3A). #1868 is the 5' primer: there is an *Nco*I site incorporated into the primer (Table 3B) and the flagellin gene starts at base 3 of the *Nco*I site. The 3' primer #1871 has a *Pst*I site incorporated downstream of the stop codon of the flagellin gene (Table 3B). PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, temperature varies (refer to Table 3A)/30'; extension, 72°C/1'; 30 cycles. The PCR product was purified using the Promega Wizard PCR purification kit (Madison WI USA) before being digested by restriction enzymes *Nco*I and *Pst*I

and cloned into the *NcoI*/*PstI* sites of plasmid pTrc99A.

Plasmid pTrc99A has a strong *trc* promoter upstream of the polylinker. Downstream of the promoter, it contains the ribosome binding site (SD sequence, see Fig 73) which is located 8bp upstream of the ATG site within the *NcoI* site. The polylinker and the SD sequence of pTrc99A are shown in Fig 73.

For flagellin genes from type strains for H30 and H32, we have the full sequence. Primer #1868 sequence is conserved in both of them. But, primer #1871 has the third base of the fourth last codon in both sequences changed from G to A to remove a *PstI* site (see Table 3B): this change did not change the amino acid coded. The expression module consists of the *trc* promoter, the SD sequence, and the full flagellin gene coding for a gene product which is same as that of the type strain. The sequence of the junction region between the SD sequence and the start of flagellin gene is shown in Fig 73.

We do not have the full sequence for the flagellin gene from the H33 type strain. In the plasmid containing the H33 type strain gene, the expression module consists of the *trc* promoter, the SD sequence, and the full flagellin gene with the first and the last 21 base pairs been determined by the primer sequences which were used for the cloning of H30 and H32. The sequence of the junction region between the SD and the start of flagellin gene is shown in Fig 73.

9.4. Flagellin genes from type strains for H4 and H17:

For the flagellin genes of H4 and H17 type strains the full sequence was not obtained, and the sequenced parts were PCR amplified and cloned into plasmid pPR1951 to give in each case a gene in which the first 26 and the last 31 codons are based on the sequence of the H7 flagellin gene of the H7 type strain.

9.4.1 Construction of expression plasmid vector

pPR1951:

The first 26 codons of the H7 flagellin gene was first PCR amplified using primers #1868 and #1872 (Table 3B). #1868 is the 5' primer: there is an *NcoI* site incorporated into the primer (Table 3B) and the flagellin gene starts at base 3 of the *NcoI* site. Primer #1872 was made to have the last two codons (codons 25 and 26) changed from CTG TCG (Leucine and Serine) to GGA TCC (Glycine and Serine) to generate a *BamHI* site. This PCR fragment was digested with *NcoI* and *BamHI* before being cloned into the *NcoI/BamHI* sites of pTrc99A to make plasmid pPR1949.

The last 31 codons (including the stop codon) of the H7 flagellin gene was PCR amplified using primers #1884 and #1871 (Table 3A). The 5' primer, #1884, has the first two of the 31 codons changed from TCG AAA (Serine and Lysine) to TCT AGA (Serine and Arginine) to generate a *XbaI* site (Table 3B). The 3' primer #1871 has a *PstI* site incorporated downstream of the stop codon (Table 3B). This PCR fragment was digested with *XbaI* and *PstI*, and then cloned into the *XbaI/PstI* sites of pPR1949 to make plasmid pPR1951.

9.4.2 *Cloning of flagellin genes from the H4 and H17 type strains:*

The central regions of flagellin genes from type strains H4 and H17 were PCR amplified using primers #1878 and #1885 (Table 3B), which have a *BamHI* and a *XbaI* incorporated at their ends respectively. PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, 65°C/30'; extension, 72°C/1'; 30 cycles. The PCR product was purified using the Promega Wizard PCR purification kit (Madison WI USA) before being digested by restriction enzymes *BamHI* and *XbaI* and cloned into the *XbaI/BamHI* sites of plasmid pPR1951 to make plasmids pPR1955 (H4) and pPR1957 (H17).

The expression module of plasmids pPR1955 and pPR1957

consists of the *trc* promoter, the SD sequence, the first 24 codons of the H7 flagellin gene (of the H7 type strain), 2 codons encoding Glycine and Serine, 292 or 293 codons of the central region based on the flagellin gene obtained from the H4 or H17 type strain respectively, 2 codons encoding Serine and Arginine, and then the last 29 codons of the H7 flagellin gene (of the H7 type strain).

10. *Expression of flagellin gene plasmids in E. coli strains lacking the fliC gene, and identification of the H antigens encoded by these plasmids:*

Plasmids carrying flagellin genes as described in section 9 (see Table 3A for a list) were electroporated into strains M2126 or P5560. Strains M2126 and P5560 do not have functional *fliC* genes, and are not motile when examined under a microscope. Transformants carrying any of the plasmids listed in Table 3A are motile when examined under a microscope. Thus, the flagellin genes in all of the plasmids are expressed.

The antigenic specificity of the flagellin of each transformant was then determined by slide agglutination.

10.1 *Flagellin genes from type strains for H2, H5, H6, H7, H9, H11, H14, H15, H18, H19, H20, H21, H23, H24, H26, H27, H28, H29, H30, H31, H32, H33, H34, H39, H41, H42, H43, H45, H46, H49, H51, H52, and H56:*

As shown in Table 3A, strains with plasmids carrying these flagellin genes expressed the same H antigen as their respective flagellin parent strain.

For flagellin specificities H2, H5, H6, H7, H9, H14, H15, H18, H19, H20, H23, H24, H26, H27, H28, H29, H31, H33, H39, H51, H52, and H56, there was no cross reaction reported between these flagellins and flagellin antisera for other H antigens [Ewing, W. H.: *Edwards and Ewing's identification of the Enterobacteriaceae.*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1986], and we conclude that we have in each case sequenced the gene

encoding the flagellin of the expected specificity from the respective type strain.

It has been observed that cross reactions exist between some type strains and certain antisera at different levels of dilution (of the antisera), being H11 with anti-H21 and anti-H40, H21 with anti-H11, H30 with anti-H32, H32 with anti-H30, H34 with anti-H24 and anti-H31, H41 with anti-H37 and anti-H39, H42 with anti-H6, H43 with anti-H37, H45 with anti-H20, H46 with anti-H17, and H49 with anti-H39 [Ewing, W. H.: Edwards and Ewing's identification of the *Enterobacteriaceae.*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1986]. We have tested strain M2126 or strain P5560 carrying plasmids containing flagellin genes obtained from each of these type strains (H11, H21, H30, H32, H34, H41, H42, H43, H45, H46, and H49) with the appropriate cross-reacting antisera.

For strain M2126 or strain P5560 carrying plasmids containing flagellin genes obtained from type strains H11, H34, H41, H42, H43, H45, H46, and H49, no cross reaction was found. We conclude that we have in each case sequenced the gene coding the flagellin of the expected specificity from the respective type strain.

Cross reaction was observed for strain P5560 carrying plasmid pPR1948 (containing the flagellin gene obtained from the H30 type strain) with anti-H32 serum, strain P5560 carrying pPR1940 (containing the flagellin gene obtained from the H32 type strain) with anti-H30 serum, and strain M2126 carrying plasmid pPR1995 (containing the flagellin gene obtained from the H21 type strain) with anti-H11 serum.

We note that the reported cross reactions between the H30 type strain and anti-H32, the H32 type strain and anti-H30, and the H21 type strain and anti-H11 happened at a higher level of dilution (of antisera) than for all other type strains with the antisera mentioned above [Ewing, W. H.: Edwards and Ewing's identification of the

Enterobacteriaceae., Elsevier Science Publishers, Amsterdam, The Netherlands, 1986]. We conclude that except for these three cases, the antiserum used were supplied at a dilution which did not exhibit cross reactions. For the three strains carrying flagellin genes cloned from type strains for H30, H32, and H21, it was necessary to further dilute the antiserum.

Strain P5560 carrying plasmid pPR1948 (containing the flagellin gene obtained from the H30 type strain) agglutinates with anti-H30 serum when the antiserum is diluted to 1:60, but agglutinates with anti-H32 serum only at a dilution of 1:10 and not at a 1:20 dilution (note that the antisera we used have been diluted before reaching our hands). In contrast, strain P5560 carrying plasmid pPR1940 (containing flagellin gene obtained from the H32 type strain) agglutinates with anti-H32 serum when the antiserum is diluted at 1:100, but agglutinates with anti-H30 serum only at a 1:10 dilution and not at a 1:10 dilution. Thus, we conclude that the flagellin genes we sequenced from type strains for H30 and H32 encode flagellins of H30 and H32 specificities respectively.

Strain M2126 carrying plasmid pPR1995 (containing the flagellin gene obtained from the H21 type strain) agglutinates with anti-H21 serum when the antiserum is diluted to 1:40, but agglutinates only with undiluted anti-H11 serum and not at a 1:10 dilution (note that the antisera we used have been diluted before reaching our hands). In contrast, strain M2126 carrying plasmid pPR1981 (containing flagellin gene obtained from the H11 type strain) did not agglutinate with anti-H21 serum. Thus, we conclude that the flagellin genes we sequenced from type strains for H21 encodes flagellin of H21 specificity.

10.2 Flagellin genes from type strains of H1 and H12:

These two genes are very similar in sequence, with 8 a.a difference between the gene products. It has been

known that some cross-reaction exists between anti-H12 serum and the H1 type strain and between anti-H1 serum and the H12 type strain [Ewing, W. H.: Edwards and Ewing's identification of the *Enterobacteriaceae.*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1986]. Strain M2126 carrying pPR1920 (carrying a flagellin gene from the H1 type strain, Table 3A) agglutinates with anti-H1 serum when the antiserum is diluted to 1:100, but agglutinates only with undiluted anti-H12 serum and not at a 1:10 dilution (please note that the antisera we used have been diluted before reaching our hands). In contrast, strain M2126 carrying plasmid pPR1990 (carrying a flagellin gene from the H12 type strain, Table 3A) agglutinates with anti-H12 serum when the antiserum is diluted at 1:100, but agglutinates only with undiluted anti-H1 serum and not at a 1:10 dilution. Thus, we conclude that the flagellin genes we sequenced from type strains for H1 and H12 encode flagellins of H1 and H12 specificities respectively.

10.3. Flagellin gene coding for H16:

Strain P5560 carrying plasmid pPR1969 agglutinated with anti-H16 serum. pPR1969 carries a flagellin gene amplified from the H3 type strain. It has been shown that this H3 type strain is a biphasic strain which can express H3 and H16 specificities [Ratiner, Y. A. (1985) "Two genetic arrangements determining flagellar antigen specificities in two diphasic *E. coli* strains. FEMS Microbiol Lett 19: 317-323]. Thus, the H3 type strain has two flagellin genes coding for H3 and H16 specificities. We conclude that we have cloned and sequenced the H16 flagellin gene from this H3 type strain.

10.4 Flagellin gene coding for H4:

The flagellin genes obtained from type strains for H4 and H17 are nearly identical, with 4 a.a. difference in the gene products. Plasmid pPR1955 carries a flagellin

gene from the H4 type strain, and plasmid pPR1957 carries a flagellin gene from the H17 type strain. Strain P5560 carrying plasmid pPR1955 or plasmid pPR1957 agglutinated with anti-H4 serum, but not with anti-H17 serum. It has been shown that the type strain for H17 is a biphasic strain which can express H17 and H4 [Ratiner, Y. A. (1985) "Two genetic arrangements determining flagellar antigen specificities in two diphasic *E. coli* strains. FEMS Microbiol Lett 19: 317-323]. The flagellin gene obtained from type strain for H44 is also highly similar to that obtained from the H4 type strain, with 2 a.a. difference in the gene products. It has been shown that the H44 type strain has two complete flagellin genes, being H4 and H44 [Ratiner, Y. A. (1998) "New flagellin specifying genes in some *E. coli* strains" J. Bacteriol 180: 979-984]. Thus, we conclude that all the three flagellin genes (obtained from type strains for H4, H17 and H44, and sequenced) encode the H4 flagellin, and that the flagellin genes for H17 and H44 specificities have not yet been cloned.

10.5 *Flagellin gene coding for H10:*

The flagellin genes obtained from type strains for H10 and H50 are nearly identical, with 3 a.a. difference in the gene products. Strain P5560 carrying plasmid pPR1923 (which carries a flagellin gene from the H10 type strain) agglutinated with anti-H10 serum. We conclude that the sequence obtained from the H10 type strain encodes the H10 flagellin. It is not clear if the sequence obtained from the H50 type strain encodes H10 or H50 (see below section for H50).

10.6 *Flagellin gene coding for H38:*

The flagellin genes obtained from type strains for H38 and H55 are nearly identical, with only 1 a.a. difference in the gene products. Strain M2126 carrying plasmid pPR1984 (carrying the flagellin gene from the type strain H38) agglutinated with anti-H38 serum, but not with

anti-H55 serum. It also has been shown that the type strain for H55 has two complete flagellin genes coding for H55 and H38 specificities [Ratiner, Y. A. (1998) "New flagellin specifying genes in some *E. coli* strains" J. Bacteriol 180: 979-984]. Thus, we conclude that both
5 cloned genes encode the H38 flagellin.

10.7 Summary:

Flagellin genes coding for 39 H antigens have been
10 identified, being those for specificities H1, H2, H4, H5, H6, H7, H9, H10, H11, H12, H14, H15, H16, H18, H19, H20, H21, H23, H24, H26, H27, H28, H29, H30, H31, H32, H33, H34, H38, H39, H41, H42, H43, H45, H46, H49, H51, H52, and H56.

11. Comparison and alignment of the flagellin genes:

Programs Pileup [Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12 (1984) 387-395] and Multicomp [Reeves, P.R., Farnell, L. and Lan, R.: MULTICOMP: a program for preparing sequence data for phylogenetic analysis. CABIOS 10 (1994) 281-284] were
20 used.

The previously published sequence of H1 (GenBank accession number L07387) was extracted from GenBank and used. Because we did not sequence H36 and H53 flagellin genes and we did not have the H16 type strain, we only compared 51 flagellin genes of H type strains and the *fliC* genes from the additional 10 H7 strains.

Among the H7 *fliC* genes, the percentage of DNA difference ranged from 0.0 to 2.39%. The flagellin genes from type strains for H40 and H8 are identical. Some others are nearly identical: H21 and H47 (1.5% difference), H12 and H1 (2.6% difference), H10 and H50 (0.3% difference), H38 and H55 (0.1% difference), H4, H44 and H17 are very similar, the pairwise difference ranging
35 from 0.33% to 0.87%.

For the flagellin genes obtained from type strains for H4, H17 and H44, we have shown that all the three genes encode flagellin with the H4 specificity (see above). For the flagellin genes obtained from type strains fro H21 and H47, and H38 and H55, we have confirmed the specificities for one for each pair and have good reason to conclude that both genes of each pair encode the same H specificity (see above section), being that for H21 and H38 specificities respectively.

For the flagellin genes obtained from type strains for H10 and H50, we have confirmed that the one from the H10 type strain encodes H10 specificity. As these two genes are highly similar, we have presumed that they both encode H10 specificity.

In the cases where the flagellin gene from two type strains is near identical, we conclude that both genes code for flagellin of the same H specificity and that one or other strain has an additional locus which carries the functional gene, although the flagellin genes sequenced do not appear to be mutated.

We have shown by cloning and expression that the flagellin genes obtained from the H1 and H12 type strains encode H1 and H12 specificities respectively (see above section). The neucleotide difference between these two genes is higher at 2.6% (see above), but still within the normal range for variation within a gene in *E. coli*. The two antigens cross react, and this cross reaction must be due to the high level similarity of the flagellins encoded by these two genes.

As discussed above, genes encoding some H antigens have been shown to be located at loci other than *fliC*. H3, H36, H47, H53 have been shown to be at a locus called *flkA*, H44 and H55 at *fllA*, and H54 at *flmA* [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984]. However, these strains may carry a *fliC* in addition to *flkA*, *fllA* or *flmA* [Ratiner Y A (1998) "New flagellin-specifying genes in some

Escherichia coli strains" J. Bacteriol. 180 979-984].

The flagellin gene encoding H48 was previously sequenced from *E. coli* strain K-12 [Ku wajima G, Asaka J, Fujiwara T, Node K and Kondo E "Nucleotide sequence of the hag gene encoding flagellin of *Escherichia coli*" J Bacteriol. 168 (1986) 1479-1483]. We have sequenced the *fliC* gene from the H48 type strain, and found that it is identical to that from K-12.

The H54 gene is known to be at *flmA* [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984]

and the finding of a non-functional presumptive *fliC* locus in the H54 strain shows that it is present but not expressed. However, we have not amplified and sequenced the functional *flmA* gene of this strain.

Using the 43 unique sequences (being the 39 identified genes with confirmed specificities and the flagellin genes obtained from the H8 (or H40), H25, H37, and H48 type strains) and the sequences from the two non-functional flagellin genes (from H type strains H35 and H54) (see Table 3) we have been able to determine antigen specific primers for each of the H antigen specificities and thereby show that it is practicable to detect *E.coli* strains carrying specific H antigens without false positives from strains of other H types. There is no reason to expect that the addition of 11 sequences to the 43 unique sequences obtained will affect the general conclusion, as unlike previous reports, our study covers flagellin sequences for a substantial majority of known *E. coli* H antigen specificities.

Our study of 11 H7 genes from strains of eight different O antigens shows limited variation which was such that the variation within genes for H antigens does not affect the ability to select antigen specific primers. O:H combinations in general define a strain and as some of the strains thus defined were quite distant from each other in a study by Whittam [Whittam T S, wolfe M L,

Wachsmuth I K, Orskov I and Wilson R A "Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea" Infect. Immun. 61 (1993) 1619-1629] the variation we observe is thought to represent that generally present in H7 genes. We also obtained more than one sequences for flagellin genes for H specificities H4, H10, and H38, and again the level of variation within a given specifities is very low. However, there is a low possibility that primers chosen without knowledge of the variation within genes of each H specificity could fail to give positive results with some isolates due to chance choice of primers which cover a base or bases which contribute to this low level variation. The variation within the H7 genes is in the normal range for variation within a gene in *E. coli* and if this possibility did occur it would be easy to use an alternate primer pair. For example, if a first primer in a primer pair is unable to hybridise to a target region because of low level variation in that region, a positive result may be achieved by using a second primer in that pair together with a third primer, whether or not the third primer is specific for the flagellin gene. Where the third primer is not specific for the flagellin gene, the specificity of the primer pair derives from the specificity of the second primer. The observation that the overall level of variation within gene for a given H specificity is very low making it extremely unlikely that the regions covered by the two primers specific for H specificity would both have undergone change in the same strain.

There are 54 known H antigens for *E. coli* and of these there are 11 H antigen specificities for which we do not as yet have sequence. It will be easy to determine these sequences and determine primer pairs specific for these H antigens by comparing these sequences with the 45 obtained sequences (see Table 3), and also modify the primers selected for any H antigen for which we already

know the sequence in the unlikely event that there is a possibility of false positives with the primers selected.

The sequences for the remaining H antigens can be obtained in one of the following ways:

5

10

15

20

25

1. where we have two bands by PCR (H36 and H53 type strains), we purify each and sequence, and also clone each into a strain mutated in its *fliC* gene and determine the H antigen expressed by use of specific sera. In this way a specific sequence can be related to an H antigen specificity. The other band which represents an H antigen gene for a different specificity is expected to include a mutant gene or a gene similar to one of those for a known H specificity, but if not may represent a new specificity for which primer pairs could be selected. It may be difficult to obtain expression of flagellin genes when cloned from *E. coli* due to cloning together with regulatory sequences which prevent expression. This is easily avoided by cloning the major segment of the gene into a functioning *fliC* gene to replace the equivalent segment of that gene, using standard site directed mutagenesis to give suitable restriction sites within the cloned gene and incorporating those restriction sites into primers used to amplify the major segment of the gene to be studied to facilitate the cloning. We have cloned and sequenced the PCR bands from the H36 and the H55 type strains using this method (see section 16).

30

35

2. Where two or more strains have the same flagellin gene sequence, the genes are cloned as above and the H antigen specificity represented by this sequence is determined. This identifies the strain in which the expected gene is expressed and also those strains for which we have sequenced a gene which is not being expressed. We then clone the gene for the antigen expressed in these strains by making a bank of plasmid clones using chromosomal DNA and select for a clone which

is expressing an H antigen different from the one represented by the known sequence. This can be done by taking advantage of the fact that the H antigen is on flagellin, the protein of the bacterial flagellum used for movement of the bacteria. In the presence of antibodies specific to that flagellum the bacteria cannot swim. For selection the clones are placed in a situation in which motile cells can swim away from the others and be collected. There are many versions of these techniques and any could be used. One version is to place the bacteria on a nutrient agar plate with reduced agar content such that bacteria can swim away from the site of inoculation. This is easily seen as growth on the plate and a sample of the bacteria which are motile can be recovered and cultivated. In this way bacteria carrying cloned H antigen genes can be selected. If the medium in the plate has antibody added to it only bacteria which express an H antigen different to that recognised by the antiserum will be able to swim. Specifically if the antiserum used is specific for the H antigen expressed by the gene for which we have sequence, only clones which express a different H antigen, such as those expressing the H antigen expressed by the H type strains used to make the plasmid, will be selected. Once the clone is obtained, the H antigen gene can be sequenced.

Our work has shown that there are at least 7 cases where the H antigen type strains carry two H antigen genes which appear to be complete and have the potential to function. However, while *E. coli* does not (in general) have a capacity to express more than one flagellin gene, it is striking that there are several loci for flagellin genes [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984]. Several of the pairs of H type strains with identical or near identical sequence do not include any of the H antigen types shown by Ratiner [Ratiner Y A

(1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] to map other than at *fliC* although these predominate. This suggests that there are additional cases where the expressed gene is not the only flagellin gene present. However the fact that many of the cases where we obtained flagellin genes of identical or near identical sequence and/or two flagellin genes from one strain involve type strains found by Ratiner [Ratiner Y A (1998) "New flagellin-specifying genes in some *Escherichia coli* strains" J. Bacteriol. 180 979-984] to map away from *fliC* are among those near identical to others, indicates that the phenomenon is of limited extent. Nonetheless it remains possible even where only one gene has been obtained by PCR, that it is one of a pair of flagellin genes, the other not being amplified by the primers used, and further that it is the one not amplified which is expressing the H antigen of the strain.

It will therefore be necessary to clone as described above each of the flagellin genes we have sequenced and confirm that it expresses the expected antigen to ensure that the invention give results corresponding to those of the traditional serotyping scheme. In the event that it does not, the gene for the type antigen can be cloned and sequenced by the means described above.

The 11 H7 *fliC* sequences fell into three groups, one comprising the genes from the O157:H7 and O55:H7 strains, which were identical, as expected given the proposed relationship between the clones. It has been shown that *E. coli* O157:H7 and O55:H7 clones are closely related [Whittam T S, wolfe M L, Wachsmuth I K, Orskov I and Wilson R A "Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea" Infect. Immun. 61 (1993) 1619-1629] thus it was expected that the H7 *fliC* genes from O157 and O55 would be identical. Among the H7 *fliC* sequences, we can identify primers specific to the H7 *fliC* gene for each of the three H7 groups. Two of these primers in combination with an H7

specific primer gave two primer pairs specific for the H7 gene of from the O157:H7 and O55:H7 clones.

5 13. *Specific oligonucleotide primers for each of the 43 flagellin genes*

Two oligonucleotide primers were chosen based on each of the 43 sequences. None of them had more than 85% identity with any other of 61 flagellin gene sequences. Thus, these primers are specific for each H type. These primers are listed in Table 3.

10 The flagellin gene of the H54 type strain is a mutated gene. It has an insertion sequence (IS1222) inserted into a normal flagellin gene of H21. Thus, primers for H21 would amplify a fragment of different size in H54. We also provide 2 primers based on the insertion sequence (see H54 row in Table 3), and the use of one of them in combination with one of the H21 primers will generate a PCR band only in H54, which will also differentiate those strain carrying the mutated H21 gene from those expressing the H21 flagellin gene.

20 The *flic* gene of H35 type strain is also a mutated gene. It has an insertion sequence (IS1) inserted into a normal flagellin gene of H11. Thus, primers for H11 would amplify a fragment of different size in H35. We also provide 2 primers based on the insertion sequence (see H35 row in Table 3), and the use of one of them in combination with one of the H11 primers will generate a PCR band only in H35, which will also differentiate those strain carrying the mutated H11 gene from those expressing the H11 flagellin gene.

30 14. *Testing of the H7 specific oligonucleotide primers*

35 Primer pair #1806/#1809 (see Table 3) was used to carry out PCR on chromosomal DNA samples of all the 54 H type strains and the H7 strains listed in Table 1. PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, 58°C/30'; extension,

72°C/1'; 30 cycles. PCR reaction was carried out in an volume of 50ul for each of the chromosomal sample. After the PCR reaction, 5µl PCR product from each sample was run on an agarose gel to check for amplified DNA.

5 Primer pairs #1806/#1809 produced a band of predicted size with all the 11 strains expressing H7, but gave no band with other H type strains. Thus, these primers are H7 specific.

10 15. *Testing of oligonucleotide primers specific to H7 of O157 and O55:*

 Based on a comparison of the *fliC* sequences of 11 different H7 strains, we have identified two oligonucleotides [#1696 (5'-GGCCTGACTCAGGCGGCC) at
15 positions 178 to 195 in M527 and #1697 (5'-GAGTTACCGGCCTGCTGA) positions 1700-1683 in M527] which are unique to H7 of O157 and O55. Although not identical to any parts of the *fliC* sequences of any other H7 strains, these two primers are identical or have high level
20 similarity to *fliC* genes of some other H types. However a combination of one of these primers with one of the H7 specific primers can give specificity for H7 of O157:H7 and O55:H7 *E. coli*.

 Primer pairs #1696/#1809 and #1697/#1806 were used to
25 carry out PCR on chromosomal DNA samples of all the H type strains and the H7 strains listed in Table 1. PCR reactions were carried out under the following conditions: denaturing, 94°C/30'; annealing, 61°C/30' (for #1696/#1809) or 60°C/30' (for #1697/#1806); extension,
30 72°C/1'; 30 cycles. PCR reaction was carried out in an volume of 50µl for each of the chromosomal samples. After the PCR reaction, 5µl PCR product from each sample was run on an agarose gel to check for amplified DNA.

 Both primer pairs produced a band of predicted size
35 with both of the O157:H7 strains (strains M1004 and M527, see Table 1), and the O55:H7 strain (strain M1686, see Table 1), but gave no band with other strains. Thus, these

two pairs of primers are specific to H7 genes of O157:H7 and O55:H7 *E. coli* strains.

16. Identification of flagellin genes for the remaining 15
5 H specificities.

16.1. Sequencing the potential *flkA* gene coding for the
H36 flagellin:

Using primers #1431 (5'- atg gca caa gtc att aat acc
10 caa c) and #1432 (5'- cta acc ctg cag cag aga ca), we have
amplified two bands from the H36 type strain. PCR reaction
was carried out under the following conditions:
denaturing, 94°C/30'; annealing, 57°C/30'; extension,
72°C/1'; 30 cycles. These two PCR fragments were then
15 cloned into the pGEM-T vector using the Promega pGEM-T
cloning kit (Madison WI USA) to make plasmids pPR1992 and
pPR1993. Inserts from both plasmids were first sequenced
using the M13 universal primers (which bind to the pGEM-T
DNA flanking the insertion site). For pPR1992, primers
20 based on the sequence obtained were then used to sequence
further, and this procedure was repeated until the insert
was fully sequenced.

The sequence of the insert of pPR1992 is identical to
that of the H12 flagellin gene sequence except perhaps for
25 the first 8 and last 7 codons which are encoded by the PCR
primers in plasmid pPR1992. We have only sequenced the
two ends of the insert of plasmid pPR1993 (Figures 71 and
72), and the sequences of the two ends of the insert of
pPR1993 are very similar to ends of other sequenced
30 flagellin genes. We conclude that the insert of plasmid
pPR1993 encodes a flagellin gene. The full sequence of
the insert of plasmid pPR1993 can be obtained using the
same method as for the sequencing of the insert of plasmid
pPR1992. It is known that *flkA* gene encodes the H36
35 flagellin [Ratiner, Y. A. (1998) "New flagellin specifying
genes in some *E. coli* strains" J. Bacteriol 180: 979-984],
and it is highly likely that plasmid pPR1993 contains the

flkA gene of the H36 type strain. H specificities can be confirmed by slide agglutination.

The currently uncharacterised sequence of both ends and of DNA flanking these two sequenced genes can be obtained by PCR walking and sequencing. Methods for PCR walking from a known sequence to an unknown region in chromosomal DNA are available (see [Siebert, P. D. , A. Chench, D. E. Kellogg, A. Lukyanov and S. A. Lukyanov (1995) "An improved PCR method for walking in uncloned genomic DNA." Nuc. Acids Res. 23: 1087-1088])).

The sequenced genes then can be PCR amplified and cloned using the method(s) described in section 9. Flagellins expressed by strain M2126 carrying these plasmids then can be determined by use of specific sera.

The sequences flanking the *flkA* gene can then be used to PCR amplify other *flkA* genes (see below).

16.2 The *flkA* genes coding for H3, H47 and H53:

It has been shown that flagellins H3, H47 and H53 are encoded by *flkA* genes in the type strains [Ratiner, Y. A. (1998) "New flagellin specifying genes in some *E. coli* strains" J. Bacteriol 180: 979-984]. These genes can be PCR amplified using primers based on the sequences flanking the *flkA* gene in the H36 type strain. These PCR fragments can then be sequenced, and the genes expressed in strain M2126 for the identification of these genes.

16.3 The *fllA* genes coding for H44 and H55:

It is known that flagellins H44 and H55 are coded by *fllA* genes.

16.3.1 The H55 flagellin gene:

Using primers #1868 and #1870 (Table 3B), we have amplified two bands from the H55 type strain. PCR reaction was carried out under the following conditions: denaturing, 94°C/30'; annealing, 50°C/30'; extension, 72°C/1'; 30 cycles. These two PCR fragments were then

cloned into the pGEM-T vector using the Promega pGEM-T cloning kit (Madison WI USA) to make plasmids pPR1994 and pPR1989. Inserts from both plasmids were first sequenced using the M13 universal primers (which bind to the pGEM-T DNA flanking the insertion site). Primers based on the sequence obtained were then used to sequence further, and this procedure was repeated until both inserts were fully or partly sequenced.

The sequence of the insert of pPR1994 is highly similar to that of the flagellin gene of the H38 type strain, with 1 amino acid difference in the gene products. We have only sequenced the two ends of the insert of plasmid pPR1989 (figures 70A and 70B), and the sequences of the two ends of the insert of pPR1989 are very similar to ends of other sequenced flagellin genes. We conclude that the insert of plasmid pPR1989 encodes a flagellin gene. The full sequence of the insert of plasmid pPR1989 can be obtained using the same method as for the sequencing of the insert of plasmid pPR1994. It is known that the H55 type strain carries flagellin genes for both H38 and H55, and that the H55 flagellin gene is at the *flaA* locus [Ratiner, Y. A. (1998) "New flagellin specifying genes in some *E. coli* strains" J. Bacteriol 180: 979-984]. Thus, it is highly likely that plasmid pPR1989 contains the *flaA* gene of the H55 type strain.

The currently uncharacterised sequence of both ends and of DNA flanking these two sequenced genes can be obtained by PCR walking and sequencing. Methods for PCR walking from a known sequence to an unknown region in chromosomal DNA are available (see [Siebert, P. D. , A. Chenchi, D. E. Kellogg, A. Lukyanov and S. A. Lukyanov (1995) "An improved PCR method for walking in uncloned genomic DNA." Nuc. Acids Res. 23: 1087-1088]).

The sequenced genes then can be PCR amplified and cloned using the method(s) described in section 9. Flagellins expressed by strain M2126 carrying these plasmids then can be determined by use of specific sera.

16.3.2 *The H44 flagellin gene:*

The sequence information for DNA flanking the *f11A* gene in the H55 type strain can then be used to PCR, sequence and identify the *f11A* gene in the H44 type strain.

16.4 *The flmA gene coding for H54:*

This gene can be cloned by making a bank of plasmid clones in strain M2126 using chromosomal DNA of the H54 type strain and selecting for a transformant which is motile on an agar plate. This is done by taking advantage of the fact that the H antigen is on flagellin, the protein of the bacterial flagellum used for movement of the bacteria. Strain M2126 lacks flagellin. Once the clone(s) is obtained and identified by use of anti-H54 serum, the flagellin gene can be sequenced. It is possible that clones expressing different flagellin specificities can be obtained, and each of them can be identified by using different sera.

16.5 *The flagellin genes obtained from the H37 and H48 type strains:*

We have used primers #1868 and #1869 (both were based on the sequence obtained from the H48 type strain, also see section 9) and primers #1868 and #1870 (both were based on the sequences of the H7 flagellin gene of the H7 type strain, also see section 9) to PCR amplify and clone the sequenced flagellin genes from the H48 and H37 type strains respectively. Strain P5560 carrying the plasmid containing either the cloned gene was not motile and did not react with the appropriate antisera. It is highly likely that mutations have occurred due to PCR errors. This can be resolved by re-amplification and re-cloning of the genes.

16.6 *The flagellin gene obtained from the H25 type*

strain:

The flagellin gene sequence we first obtained from the H25 type strain lacks 23 and 21 codons at 5' and 3' ends respectively. We could not amplify the full gene from the H25 type strain using primers based on the H7 flagellin gene of the H7 type strain, and it was necessary to get the full sequence of this flagellin gene by other means.

We have used primers (#2650: 5' - cag cga tga aat act tgc cat and #2648: 5' - caa tgc ttc gtg acg cac) based on the genes (*fliD* and *fliA* respectively) flanking *fliC* gene in *E. coli* K-12 [Blattner, F. R., G. I. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley and et al. (1997) "The complete genome sequence of *E. Coli* K12" Science 277: 1453-1474] and primers (#2658: 5' - gcc tga gtc aga cct ttg and # 2653 5' - aac ctg tct gaa gcg cag) based on the flagellin sequence obtained from the H25 type strain to PCR amplify both ends of the flagellin gene. The PCR product was then sequenced, and we have now obtained the full flagellin gene sequence and sequence for the DNA flanking the flagellin gene from type strain H25 (Figure 69). Now, it is straightforward to PCR amplify, clone and express, and identify this gene using the methods described in sections 9 and 10.

16.7 *The flagellin genes obtained from the H8 and H40 type strains:*

The flagellin gene sequences obtained from both the H8 and H40 type strains lack 18 and 15 codons at 5' and 3' ends respectively. We have used primers based on the H7 flagellin gene of the H7 type strain to PCR amplify and clone the full genes from these two strains. Strain M2126 carrying plasmid made this way was not motile under microscope and did not react with the appropriate antisera. This could be due to PCR errors as mentioned in section 16.5 or perhaps the first and last few amino acids encoded by the primers (based on H7 flagellin gene) are

uncompatible in this case.

The full sequence of the full gene can be obtained using method described in section 16.6. The flagellin gene can then be PCR amplified, cloned and expressed, and identified using the methods described in sections 9 and 10.

The gene products of the flagellin genes obtained from the H8 and H40 type strains are identical. Thus, one of these two H specificities must be encoded by a unknown gene, and it can be cloned and identified using the method described in the section 16.8.

16.8 *Flagellin genes coding for H17, H35, and H50:*

As mentioned above, the sequenced flagellin genes from the H17 and H50 type strains encode H4 and H10 specificities respectively. The flagellin gene sequence obtained from the H35 strain has a insertion and encodes a non-functional gene (see section 8). Thus, genes coding for these flagellins have not been identified, and their location is unknown. One can use primers based on DNA flanking *fliC*, *fliA*, *flkA*, and *flmA* to do PCR on the type strain for each of the flagellin antigen. PCR products can then be sequenced, and possible genes can be cloned, expressed and identified then.

If the target gene is not PCR amplified using primers based on sequence of these loci or sequence flanking these loci, it can be cloned by making a bank of plasmid clones in strain M2126 using chromosomal DNA of the type strain and selecting for a transformant which is motile on an agar plate. This is done by taking advantage of the fact that the H antigen is on flagellin, the protein of the bacterial flagellum used for movement of the bacteria. Strain M2126 lacks flagellin. Once the clone(s) is obtained and identified by use of antisera, the flagellin gene can be sequenced. It is possible that clones expressing different flagellin antigens can be obtained,

and each of them can be identified by using different antisera. Antiserum for H50 can be prepared using standard methods [Ewing, W.H.: Edwards and Ewing's identification of the *Enterobacteriaceae.*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1986].

O antigen

Materials and Methods-part 1

The experimental procedures for the isolation and characterisation of the *E. coli* O111 O antigen gene cluster (position 3,021-9,981) are according to Bastin D.A., et al. 1991 "Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5:9 2223-2231 and Bastin D.A. and Reeves, P.R. 1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". *Gene* 164: 17-23.

A. Bacterial strains and growth media

Bacteria were grown in Luria broth supplemented as required.

B. Cosmids and phage

Cosmids in the host strain x2819 were repackaged *in vivo*. Cells were grown in 250mL flasks containing 30mL of culture, with moderate shaking at 30°C to an optical density of 0.3 at 580 nm. The defective lambda prophage was induced by heating in a water bath at 45°C for 15min followed by an incubation at 37°C with vigorous shaking for 2hr. Cells were then lysed by the addition of 0.3mL chloroform and shaking for a further 10min. Cell debris were removed from 1mL of lysate by a 5min spin in a microcentrifuge, and the supernatant removed to a fresh microfuge tube. One drop of chloroform was added then shaken vigorously through the tube contents.

C. DNA preparation

Chromosomal DNA was prepared from bacteria grown overnight at 37°C in a volume of 30mL of Luria broth. After harvesting by centrifugation, cells were washed and

resuspended in 10mL of 50mMTris-HCl pH 8.0. EDTA was added and the mixture incubated for 20min. Then lysozyme was added and incubation continued for a further 10min. Proteinase K, SDS, and ribonuclease were then added and the mixture incubated for up to 2hr for lysis to occur. All incubations were at 37°C. The mixture was then heated to 65°C and extracted once with 8mL of phenol at the same temperature. The mixture was extracted once with 5mL of phenol/chloroform/iso-amyl alcohol at 4°C.

Residual phenol was removed by two ether extractions. DNA was precipitated with 2 vols. of ethanol at 4°C, spooled and washed in 70% ethanol, resuspended in 1-2mL of TE and dialysed. Plasmid and cosmid DNA was prepared by a modification of the Birnboim and Doly method [Birnboim, H. C. and Doly, J. (1979) "A rapid alkaline extraction procedure for screening recombinant plasmid DNA" *Nucl. Acid Res.* 7:1513-1523]. The volume of culture was 10mL and the lysate was extracted with phenol/chloroform/iso-amyl alcohol before precipitation with isopropanol. Plasmid DNA to be used as vector was isolated on a continuous caesium chloride gradient following alkaline lysis of cells grown in 1L of culture.

D. Enzymes and buffers.

Restriction endonucleases and DNA T4 ligase were purchased from Boehringer Mannheim (Castle Hill, NSW, Australia) or Pharmacia LKB (Melbourne, VIC Australia). Restriction enzymes were used in the recommended commercial buffer.

E. Construction of a gene bank.

Individual aliquots of M92 chromosomal DNA (strain Stoke W, from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen S, Denmark) were partially digested with 0.2U *Sau3A1* for 1-15mins. Aliquots giving the greatest proportion of fragments in the size range of approximately 40-50kb were selected and ligated to vector pPR691 previously digested with *BamH1* and *PvuII*. Ligation mixtures were packaged *in vitro* with packaging extract.

The host strain for transduction was x2819 and recombinants were selected with kanamycin.

F. Serological procedures.

Colonies were screened for the presence of the 0111 antigen by immunoblotting. Colonies were grown overnight, up to 100 per plate then transferred to nitrocellulose discs and lysed with 0.5N HCl. Tween 20 was added to TBS at 0.05% final concentration for blocking, incubating and washing steps. Primary antibody was *E. coli* O group 111 antiserum, diluted 1:800. The secondary antibody was goat anti-rabbit IgG labelled with horseradish peroxidase diluted 1:5000. The staining substrate was 4-chloro-1-naphthol. Slide agglutination was performed according to the standard procedure.

G. Recombinant DNA methods.

Restriction mapping was based on a combination of standard methods including single and double digests and sub-cloning. Deletion derivatives of entire cosmids were produced as follows: aliquots of 1.8mg of cosmid DNA were digested in a volume of 20ml with 0.25U of restriction enzyme for 5-80min. One half of each aliquot was used to check the degree of digestion on an agarose gel. The sample which appeared to give a representative range of fragments was ligated at 4°C overnight and transformed by the CaCl_2 method into JM109. Selected plasmids were transformed into sf174 by the same method. P4657 was transformed with pPR1244 by electroporation.

H. DNA hybridisation

Probe DNA was extracted from agarose gels by electroelution and was nick-translated using [α - ^{32}P]-dCTP. Chromosomal or plasmid DNA was electrophoresed in 0.8% agarose and transferred to a nitrocellulose membrane. The hybridisation and pre-hybridisation buffers contained either 30% or 50% formamide for low and high stringency probing respectively. Incubation temperatures were 42°C and 37°C for pre-hybridisation and hybridisation respectively. Low stringency washing of

filters consisted of 3 x 20min washes in 2 x SSC and 0.1% SDS. High-stringency washing consisted of 3 x 5min washes in 2 x SSC and 0.1% SDS at room temperature, a 1hr wash in 1 x SSC and 0.1% SDS at 58°C and 15min wash in 0.1 x SSC and 0.1% SDS at 58°C.

I. Nucleotide sequencing of *E. coli* O111 O antigen gene cluster (position 3,021-9,981)

Nucleotide sequencing was performed using an ABI 373 automated sequencer (CA, USA). The region between map positions 3.30 and 7.90 was sequenced using uni-directional exonuclease III digestion of deletion families made in PT7T3190 from clones pPR1270 and pPR1272. Gaps were filled largely by cloning of selected fragments into M13mp18 or M13mp19. The region from map positions 7.90-10.2 was sequenced from restriction fragments in M13mp18 or M13mp19. Remaining gaps in both the regions were filled by priming from synthetic oligonucleotides complementary to determined positions along the sequence, using a single stranded DNA template in M13 or phagemid. The oligonucleotides were designed after analysing the adjacent sequence. All sequencing was performed by the chain termination method. Sequences were aligned using SAP [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing". *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231]. The program NIP [Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961] was used to find open reading frames and translate them into proteins.

J. Isolation of clones carrying *E. coli* O111 O antigen gene cluster

The *E. coli* O antigen gene cluster was isolated according to the method of Bastin D.A., et al. [1991 "Molecular cloning and expression in *Escherichia coli* K-

12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain". *Mol. Microbiol.* 5(9), 2223-2231]. Cosmid gene banks of M92 chromosomal DNA were established in the *in vivo* packaging strain x2819. From the genomic bank, 3.3×10^3 colonies were screened with *E. coli* O111 antiserum using an immuno-blotting procedure: 5 colonies (pPR1054, pPR1055, pPR1056, pPR1058 and pPR1287) were positive. The cosmids from these strains were packaged *in vivo* into lambda particles and transduced into the *E. coli* deletion mutant Sf174 which lacks all O antigen genes. In this host strain, all plasmids gave positive agglutination with O111 antiserum.

An *Eco* R1 restriction map of the 5 independent cosmids showed that they have a region of approximately 11.5 kb in common (Figure 1). Cosmid pPR1058 included sufficient flanking DNA to identify several chromosomal markers linked to O antigen gene cluster and was selected for analysis of the O antigen gene cluster region.

K. Restriction mapping of cosmid pPR1058

Cosmid pPR1058 was mapped in two stages. A preliminary map was constructed first, and then the region between map positions 0.00 and 23.10 was mapped in detail, since it was shown to be sufficient for O111 antigen expression. Restriction sites for both stages are shown in Figure 2. The region common to the five cosmid clones was between map positions 1.35 and 12.95 of pPR1058.

To locate the O antigen gene cluster within pPR1058, pPR1058 cosmid was probed with DNA probes covering O antigen gene cluster flanking regions from *S. enterica* LT2 and *E. coli* K-12. Capsular polysaccharide (*cps*) genes lie upstream of O antigen gene cluster while the gluconate dehydrogenase (*gnd*) gene and the histidine (*his*) operon are downstream, the latter being further from the O antigen gene cluster. The probes used were pPR472 (3.35kb), carrying the *gnd* gene of LT2, pPR685 (5.3kb) carrying two genes of the *cps* cluster, *cpsB* and

cpsG of LT2, and K350 (16.5kb) carrying all of the *his* operon of K-12. Probes hybridised as follows: pPR472 hybridised to 1.55kb and 3.5 kb (including 2.7 kb of vector) fragments of *Pst*I and *Hind*III double digests of pPR1246 (a *Hind*III/*Eco*R1 subclone derived from pPR1058, Figure 2), which could be located at map positions 12.95-15.1; pPR685 hybridised to a 4.4 kb *Eco*R1 fragment of pPR1058 (including 1.3 kb of vector) located at map position 0.00-3.05; and K350 hybridised with a 32kb *Eco*R1 fragment of pPR1058 (including 4.0kb of vector), located at map position 17.30-45.90. Subclones containing the presumed *gnd* region complemented a *gnd*⁻*edd*⁻ strain GB23152. On gluconate bromothymol blue plates, pPR1244 and pPR1292 in this host strain gave the green colonies expected of a *gnd*⁺*edd*⁻ genotype. The *his*⁺ phenotype was restored by plasmid pPR1058 in the *his* deletion strain Sf174 on minimal medium plates, showing that the plasmid carries the entire *his* operon.

It is likely that the O antigen gene cluster region lies between *gnd* and *cps*, as in other *E. coli* and *S. enterica* strains, and hence between the approximate map positions 3.05 and 12.95. To confirm this, deletion derivatives of pPR1058 were made as follows: first, pPR1058 was partially digested with *Hind*III and self ligated. Transformants were selected for kanamycin resistance and screened for expression of O111 antigen. Two colonies gave a positive reaction. *Eco*R1 digestion showed that the two colonies hosted identical plasmids, one of which was designated pPR1230, with an insert which extended from map positions 0.00 to 23.10. Second pPR1058 was digested with *Sal*I and partially digested with *Xho*I and the compatible ends were re-ligated. Transformants were selected with kanamycin and screened for O111 antigen expression. Plasmid DNA of 8 positively reacting clones was checked using *Eco*R1 and *Xho*I digestion and appeared to be identical. The cosmid of one was designated pPR1231. The insert of pPR1231

contained the DNA region between map positions 0.00 and 15.10. Third, pPR1231 was partially digested with *Xho*I, self-ligated, and transformants selected on spectinomycin/ streptomycin plates. Clones were screened for kanamycin sensitivity and of 10 selected, all had the DNA region from the *Xho*I site in the vector to the *Xho*I site at position 4.00 deleted. These clones did not express the O111 antigen, showing that the *Xho*I site at position 4.00 is within the O antigen gene cluster. One clone was selected and named pPR1288. Plasmids pPR1230, pPR1231, and pPR1288 are shown in Figure 2.

L. Analysis of the E. coli O111 O antigen gene cluster (position 3,021-9,981) nucleotide sequence data

Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111". Gene 164: 17-23] partially characterised the *E. coli* O111 O antigen gene cluster by sequencing a fragment from map position 3,021-9,981. Figure 3 shows the gene organisation of position 3,021-9,981 of *E. coli* O111 O antigen gene cluster. *orf3* and *orf6* have high level amino acid identity with *wcaH* and *wcaG* (46.3% and 37.2% respectively), and are likely to be similar in function to sugar biosynthetic pathway genes in the *E. coli* K-12 colanic gene cluster. *orf4* and *orf5* show high levels of amino acid homology to *manC* and *manB* genes respectively. *orf7* shows high level homology with *rfbH* which is an abequose pathway gene. *orf8* encodes a protein with 12 transmembrane segments and has similarity in secondary structure to other *wzx* genes and is likely therefore to be the O antigen flippase gene.

Materials and Methods-part 2

A. Nucleotide sequencing of 1 to 3,020 and 9,982 to 14,516 of the *E. coli* O111 O antigen gene cluster

The sub clones which contained novel nucleotide sequences, pPR1231 (map position 0 and 1,510), pPR1237 (map position -300 to 2,744), pPR1239 (map position 2,744

to 4,168), pPR1245 (map position 9,736 to 12,007) and pPR1246 (map position 12,007 to 15,300) (Figure 2), were characterised as follows: the distal ends of the inserts of pPR1237, pPR1239 and pPR1245 were sequenced using the

5 M13 forward and reverse primers located in the vector. PCR walking was carried out to sequence further into each insert using primers based on the sequence data and the primers were tagged with M13 forward or reverse primer sequences for sequencing. This PCR walking procedure was repeated until the entire insert was sequenced. pPR1246

10 was characterised from position 12,007 to 14,516. The DNA of these sub clones was sequenced in both directions. The sequencing reactions were performed using the dideoxy termination method and thermocycling and reaction products were analysed using fluorescent dye and an ABI automated sequencer (CA, USA).

15 B. Analysis of the *E. coli* O111 O antigen gene cluster (positions 1 to 3,020 and 9,982 to 14,516 of Figure 5) nucleotide sequence data

20 The gene organisation of regions of *E. coli* O111 O antigen gene cluster which were not characterised by Bastin and Reeves [1995 "Sequence and analysis of the O antigen gene(*rfb*)cluster of *Escherichia coli* O111." Gene 164: 17-23], (positions 1 to 3,020 and 9,982 to 14,516) is shown in Figure 3. There are two open reading frames in region 1. Four open reading frames are predicted in region 2. The position of each gene is listed in Table 9.

25 The deduced amino acid sequence of *orf1* (*wbdH*) shares about 64% similarity with that of the *rfp* gene of *Shigella dysenteriae*. Rfp and WbdH have very similar hydrophobicity plots and both have a very convincing predicted transmembrane segment in a corresponding position. *rfp* is a galactosyl transferase involved in the synthesis of LPS core, thus *wbdH* is likely to be a galactosyl transferase gene. *orf2* has 85.7% identity at amino acid level to the *gmd* gene identified in the *E.*

coli K-12 colanic acid gene cluster and is likely to be a *gmd* gene. *orf9* encodes a protein with 10 predicted transmembrane segments and a large cytoplasmic loop. This inner membrane topology is a characteristic feature of all known O antigen polymerases thus it is likely that *orf9* encodes an O antigen polymerase gene, *wzy*. *orf10* (*wbdL*) has a deduced amino acid sequence with low homology with *Lsi2* of *Neisseria gonorrhoeae*. *Lsi2* is responsible for adding GlcNAc to galactose in the synthesis of lipooligosaccharide. Thus it is likely that *wbdL* is either a colitose or glucose transferase gene. *orf11* (*wbdM*) shares high level nucleotide and amino acid similarity with TrsE of *Yersinia enterocolitica*. TrsE is a putative sugar transferase thus it is likely that *wbdM* encodes the colitose or glucose transferase.

In summary three putative transferase genes and an O antigen polymerase gene were identified at map position 1 to 3,020 and 9,982 to 14,516 of *E. coli* O111 O antigen gene cluster. A search of GenBank has shown that there are no genes with significant similarity at the nucleotide sequence level for two of the three putative transferase genes or the polymerase gene. Figure 5 provides the nucleotide sequence of the O111 antigen gene cluster.

Materials and Methods-part 3

A. PCR amplification of O157 antigen gene cluster from an *E. coli* O157:H7 strain (Strain C664-1992, from Statens Serum Institut, 5 Artillerivej, 2300, Copenhagen S, Denmark)

E. coli O157 O antigen gene cluster was amplified by using long PCR [Cheng et al. 1994, "Effective amplification of long targets from cloned inserts and human and genomic DNA" P.N.A.S. USA 91: 5695-569] with one primer (primer #412: att ggt agc tgt aag cca agg gcg gta gcg t) based on the JumpStart sequence usually found in the promoter region of O antigen gene clusters [Hobbs,

et al. 1994 "The JumpStart sequence: a 39 bp element common to several polysaccharide gene clusters" Mol. Microbiol. 12: 855-856], and another primer #482 (cac tgc cat acc gac gac gcc gat ctg ttg ctt gg) based on the *gnd* gene usually found downstream of the O antigen gene cluster. Long PCR was carried out using the Expand Long Template PCR System from Boehringer Mannheim (Castle Hill NSW Australia), and products, 14 kb in length, from several reactions were combined and purified using the Promega Wizard PCR preps DNA purification System (Madison WI USA). The PCR product was then extracted with phenol and twice with ether, precipitated with 70% ethanol, and resuspended in 40mL of water.

B. Construction of a random DNase I bank:

Two aliquots containing about 150ng of DNA each were subjected to DNase I digestion using the Novagen DNase I Shotgun Cleavage (Madison WI USA) with a modified protocol as described. Each aliquot was diluted into 45ml of 0.05M Tris -HCl (pH7.5), 0.05mg/mL BSA and 10mM MnCl₂. 5mL of 1:3000 or 1:4500 dilution of DNaseI (Novagen) (Madison WI USA) in the same buffer was added into each tube respectively and 10ml of stop buffer (100mM EDTA), 30% glycerol, 0.5% Orange G, 0.075% xylene and cyanol (Novagen) (Madison WI USA) was added after incubation at 15°C for 5 min. The DNA from the two DNaseI reaction tubes were then combined and fractionated on a 0.8% LMT agarose gel, and the gel segment with DNA of about 1kb in size (about 1.5mL agarose) was excised. DNA was extracted from agarose using Promega Wizard PCR Preps DNA Purification (Madison WI USA) and resuspended in 200 mL water, before being extracted with phenol and twice with ether, and precipitated. The DNA was then resuspended in 17.25 mL water and subjected to T4 DNA polymerase repair and single dA tailing using the Novagen Single dA Tailing Kit (Madison WI USA). The reaction product (85ml containing about 8ng DNA) was then extracted with chloroform:isoamyl alcohol (24:1) once and

ligated to 3×10^{-3} pmol pGEM-T (Promega) (Madison WI USA) in a total volume of 100mL. Ligation was carried out overnight at 4°C and the ligated DNA was precipitated and resuspended in 20mL water before being electroporated into *E. coli* strain JM109 and plated out on BCIG-IPTG plates to give a bank.

C. Sequencing

DNA templates from clones of the bank were prepared for sequencing using the 96-well format plasmid DNA miniprep kit from Advanced Genetic Technologies Corp (Gaithersburg MD USA). The inserts of these clones were sequenced from one or both ends using the standard M13 sequencing primer sites located in the pGEM-T vector. Sequencing was carried out on an ABI377 automated sequencer (CA USA) as described above, after carrying out the sequencing reaction on an ABI Catalyst (CA USA). Sequence gaps and areas of inadequate coverage were PCR amplified directly from O157 chromosomal DNA using primers based on the already obtained sequencing data and sequenced using the standard M13 sequencing primer sites attached to the PCR primers.

D. Analysis of the *E. coli* O157 O antigen gene cluster nucleotide sequence data

Sequence data were processed and analysed using the Staden programs [Staden, R., 1982 "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing." *Nuc. Acid Res.* 10: 4731-4751; Staden, R., 1986 "The current status and portability of our sequence handling software". *Nuc. Acid Res.* 14: 217-231; Staden, R. 1982 "An interactive graphics program for comparing and aligning nucleic acid and amino acid sequence". *Nuc. Acid Res.* 10: 2951-2961].

Figure 4 shows the structure of *E. coli* O157 O antigen gene cluster. Twelve open reading frames were predicted from the sequence data, and the nucleotide and amino acid sequences of all these genes were then used to search the GenBank database for indication of possible function and

specificity of these genes. The position of each gene is listed in Table 9. The nucleotide sequence is presented in Figure 6.

5 *orfs* 10 and 11 showed high level identity to *manC*
and *manB* and were named *manC* and *manB* respectively. *orf7*
showed 89% identity (at amino acid level) to the *gmd* gene
of the *E. coli* colanic acid capsule gene cluster
(Stevenson G., K. et al. 1996 "Organisation of the
10 *Escherichia coli* K-12 gene cluster responsible for
production of the extracellular polysaccharide colanic
acid". J. Bacteriol. 178:4885-4893) and was named *gmd*.
orf8 showed 79% and 69% identity (at amino acid level)
respectively to *wcaG* of the *E. coli* colanic acid capsule
gene cluster and to *wbcJ* (*orf14.8*) gene of the *Yersinia*
15 *enterocolitica* O8 O antigen gene cluster (Zhang, L. et
al. 1997 "Molecular and chemical characterization of the
lipopolysaccharide O-antigen and its role in the
virulence of *Y. enterocolitica* serotype O8". Mol.
Microbiol. 23:63-76). Colanic acid and the *Yersinia* O8 O
20 antigen both contain fucose as does the O157 O antigen.
There are two enzymatic steps required for GDP-L-fucose
synthesis from GDP-4-keto-6-deoxy-D-mannose, the product
of the *gmd* gene product. However, it has been shown
recently (Tonetti, M et al. 1996 Synthesis of GDP-L-
25 fucose by the human FX protein J. Biol. Chem. 271:27274-
27279) that the human FX protein has "significant
homology" with the *wcaG* gene (referred to as *Yefb* in that
paper), and that the FX protein carries out both
reactions to convert GDP-4-keto-6-deoxy-D-mannose to GDP-
30 L-fucose. We believe that this makes a very strong case
for *orf8* carrying out these two steps and propose to name
the gene *fcl*. In support of the one enzyme carrying out
both functions is the observation that there are no genes
other than *manB*, *manC*, *gmd* and *fcl* with similar levels of
35 similarity between the three bacterial gene clusters for
fucose containing structures.

orf5 is very similar to *wbeE* (*rfbE*) of *Vibrio*

cholerae O1, which is thought to be the perosamine synthetase, which converts GDP-4-keto-6-deoxy-D-mannose to GDP-perosamine (Stroeher, U.H et al. 1995 "A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of *Vibrio cholerae*" O1. Gene 166: 33-42). *V. cholerae* O1 and *E. coli* O157 O antigens contain perosamine and N-acetyl-perosamine respectively. The *V. cholerae* O1 *manA*, *manB*, *gmd* and *wbeE* genes are the only genes of the *V. cholerae* O1 gene cluster with significant similarity to genes of the *E. coli* O157 gene cluster and we believe that our observations both confirm the prediction made for the function of *wbe* of *V. cholerae*, and show that *orf5* of the O157 gene cluster encodes GDP-perosamine synthetase. *orf5* is therefore named *per*. *orf5* plus about 100bp of the upstream region (position 4022-5308) was previously sequenced by Bilge, S.S. et al. [1996 "Role of the *Escherichia coli* O157-H7 O side chain in adherence and analysis of an *rfb* locus". *Infect. Immun.* 64:4795-4801].

orf12 shows high level similarity to the conserved region of about 50 amino acids of various members of an acetyltransferase family (Lin, W., et al. 1994 "Sequence analysis and molecular characterisation of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*". *J. Bacteriol.* 176: 7005-7016) and we believe it is the N-acetyltransferase to convert GDP-perosamine to GDP-perNAc. *orf12* has been named *wbdR*.

The genes *manB*, *manC*, *gmd*, *fcl*, *per* and *wbdR* account for all of the expected biosynthetic pathway genes of the O157 gene cluster.

The remaining biosynthetic step(s) required are for synthesis of UDP-GalNAc from UDP-Glc. It has been proposed (Zhang, L., et al. 1997 "Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O8". *Mol. Microbiol.* 23:63-76) that in *Yersinia enterocolitica* UDP-GalNAc is synthesised from UDP-GlcNAc

by a homologue of galactose epimerase (GalE), for which there is a *galE* like gene in the *Yersinia enterocolitica* O8 gene cluster. In the case of O157 there is no *galE* homologue in the gene cluster and it is not clear how UDP-GalNAC is synthesised. It is possible that the galactose epimerase encoded by the *galE* gene in the *gal* operon, can carry out conversion of UDP-GlcNAC to UDP-GalNAC in addition to conversion of UDP-Glc to UDP-Gal. There do not appear to be any gene(s) responsible for UDP-GalNAC synthesis in the O157 gene cluster.

orf4 shows similarity to many *wzx* genes and is named *wzx* and *orf2* which shows similarity of secondary structure in the predicted protein to other *wzy* genes and is for that reason named *wzy*.

The *orf1*, *orf3* and *orf6* gene products all have characteristics of transferases, and have been named *wbdN*, *wbdO* and *wbdP* respectively. The O157 O antigen has 4 sugars and 4 transferases are expected. The first transferase to act would put a sugar phosphate onto undecaprenol phosphate. The two transferases known to perform this function, WbaP (RfbP) and WecA (Rfe) transfer galactose phosphate and N-acetyl-glucosamine phosphate respectively to undecaprenol phosphate. Neither of these sugars is present in the O157 structure.

Further, none of the presumptive transferases in the O157 gene cluster has the transmembrane segments found in WecA and WbaP which transfer a sugar phosphate to undecaprenol phosphate and expected for any protein which transferred a sugar to undecaprenol phosphate which is embedded within the membrane.

The WecA gene which transfers GlcNAC-P to undecaprenol phosphate is located in the Enterobactereal Common Antigen (ECA) gene cluster and it functions in ECA synthesis in most and perhaps all *E. coli* strains, and also in O antigen synthesis for those strains which have GlcNAC as the first sugar in the O unit.

It appears that WecA acts as the transferase for

addition of GalNAc-1-P to undecaprenol phosphate for the *Yersinia enterocolitica* O8 O antigen [Zhang et al.1997 "Molecular and chemical characterisation of the lipopolysaccharide O antigen and its role in the virulence of *Yersinia enterocolitica* serotype O8" Mol. Microbiol. 23: 63-76.] and perhaps does so here as the O157 structure includes GalNAc. *WecA* has also been reported to add Glucose-1-P phosphate to undecaprenol phosphate in *E. coli* O8 and O9 strains, and an alternative possibility for transfer of the first sugar to undecaprenol phosphate is *WecA* mediated transfer of glucose, as there is a glucose residue in the O157 O antigen. In either case the requisite number of transferase genes are present if GalNAc or Glc is transferred by *WecA* and the side chain Glc is transferred by a transferase outside of the O antigen gene cluster.

orf9 shows high level similarity (44% identity at amino acid level, same length) with *wcaH* gene of the *E. coli* colanic acid capsule gene cluster. The function of this gene is unknown, and we give *orf9* the name *wbdQ*.

The DNA between *manB* and *wdbR* has strong sequence similarity to one of the H-repeat units of *E. coli* K12. Both of the inverted repeat sequences flanking this region are still recognisable, each with two of the 11 bases being changed. The H-repeat associated protein encoding gene located within this region has a 267 base deletion and mutations in various positions. It seems that the H-repeat unit has been associated with this gene cluster for a long period of time since it translocated to the gene cluster, perhaps playing a role in assembly of the gene cluster as has been proposed in other cases.

Materials and Methods - part 4

To test our hypothesis that O antigen genes for transferases and the *wzx*, *wzy* genes were more specific than pathway genes for diagnostic PCR, we first carried out PCR using primers for all the *E. coli* O16 O antigen

genes (Table 7). The PCR was then carried out using PCR primers for *E.coli* 0111 transferase, *wzx* and *wzy* genes (Table 8, 8A). PCR was also carried out using PCR primers for the *E. coli* 0157 transferase, *wzx* and *wzy* genes (Table 9, 9A).

Chromosomal DNA from the 166 serotypes of *E. coli* available from Statens Serum Institut, 5 Artillerivej, 2300 Copenhagen Denmark was isolated using the Promega Genomic (Madison WI USA) isolation kit. Note that 164 of the serogroups are described by Ewing W. H.: Edwards and Ewings "Identification of the Enterobacteriaceae" Elsevier, Amsterdam 1986 and that they are numbered 1-171 with numbers 31, 47, 67, 72, 93, 94 and 122 no longer valid. Of the two serogroup 19 strains we used 19ab strain F8188-41. Lior H. 1994 ["Classification of *Escherichia coli* In *Escherichia coli* in domestic animals and humans pp 31-72. Edited by C.L. Gyles CAB international] adds two more numbered 172 and 173 to give the 166 serogroups used. Pools containing 5 to 8 samples of DNA per pool were made. Pool numbers 1 to 19 (Table 4) were used in the *E. coli* 0111 and 0157 assay. Pool numbers 20 to 28 were also used in the 0111 assay, and pool numbers 22 to 24 contained *E. coli* 0111 DNA and were used as positive controls (Table 5). Pool numbers 29 to 42 were also used in the 0157 assay, and pool numbers 31 to 36 contained *E. coli* 0157 DNA, and were used as positive controls (Table 6). Pool numbers 2 to 20, 30, 43 and 44 were used in the *E. coli* 016 assay (Tables 4 to 6). Pool number 44 contained DNA of *E. coli* K-12 strains C600 and WG1 and was used as a positive control as between them they have all of the *E. coli* K-12 016 O antigen genes.

PCR reactions were carried out under the following conditions: denaturing 94°C/30"; annealing, temperature varies (refer to Tables)/30"; extension, 72°C/1'; 30 cycles. PCR reaction was carried out in a volume of 25mL for each pool. After the PCR reaction, 10mL PCR

product from each pool was run on an agarose gel to check for amplified DNA.

Each *E. coli* chromosomal DNA sample was checked by gel electrophoresis for the presence of chromosomal DNA and by PCR amplification of the *E. coli mdh* gene using oligonucleotides based on *E. coli* K-12 [Boyd et al. (1994) "Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*" Proc. Nat. Acad. Sci. USA. 91:1280-1284.] Chromosomal DNA samples from other bacteria were only checked by gel electrophoresis of chromosomal DNA.

A. Primers based on *E. coli* O16 O antigen gene cluster sequence.

The O antigen gene cluster of *E. coli* O16 was the only typical *E. coli* O antigen gene cluster that had been fully sequenced prior to that of O111, and we chose it for testing our hypothesis. One pair of primers for each gene was tested against pools 2 to 20, 30 and 43 of *E. coli* chromosomal DNA. The primers, annealing temperatures and functional information for each gene are listed in Table 8.

For the five pathway genes, there were 17/21, 13/21, 0/21, 0/21, 0/21 positive pools for *rmlB*, *rmlD*, *rmlA*, *rmlC* and *glf* respectively (Table 7). For the *wzx*, *wzy* and three transferase genes there were no positives amongst the 21 pools of *E. coli* chromosomal DNA tested (Table 7). In each case the #44 pool gave a positive result.

B. Primers based on the *E. coli* O111 O antigen gene cluster sequence.

One to four pairs of primers for each of the transferase, *wzx* and *wzy* genes of O111 were tested against the pools 1 to 21 of *E. coli* chromosomal DNA (Table 8). For *wbdH*, four pairs of primers, which bind

to various regions of this gene, were tested and found to be specific for O111 as there was no amplified DNA of the correct size in any of those 21 pools of *E. coli* chromosomal DNA tested. Three pairs of primers for *wbdM* were tested, and they are all specific although primers #985/#986 produced a band of the wrong size from one pool. Three pairs of primers for *wzx* were tested and they all were specific. Two pairs of primers were tested for *wzy*, both are specific although #980/#983 gave a band of the wrong size in all pools. One pair of primers for *wbdL* was tested and found unspecific and therefore no further test was carried out. Thus, *wzx*, *wzy* and two of the three transferase genes are highly specific to O111.

Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in *E. coli*. The primers, annealing temperatures and positions for each gene are in Table 8.

The O111 assay was also performed using pools including DNA from O antigen expressing *Yersinia pseudotuberculosis*, *Shigella boydii* and *Salmonella enterica* strains (Table 8A). None of the oligonucleotides derived from *wbdH*, *wzx*, *wzy* or *wbdM* gave amplified DNA of the correct size with these pools. Notably, pool number 25 includes *S. enterica* Adelaide which has the same O antigen as *E. coli* O111: this pool did not give a positive PCR result for any primers tested indicating that these genes are highly specific for *E. coli* O111.

Each of the 12 pairs binding to *wbdH*, *wzx*, *wzy* and *wbdM* produces a band of predicted size with the pools containing O111 DNA (pools number 22 to 24). As pools 22 to 24 included DNA from all strains present in pool 21 plus O111 strain DNA (Table 5), we conclude that the 12 pairs of primers all give a positive PCR test with each of three unrelated O111 strains but not with any other strains tested. Thus these genes are highly specific for *E. coli* O111.

C. Primers based on the *E. coli* 0157 O antigen gene cluster sequence.

Two or three primer pairs for each of the transferase, *wzx* and *wzy* genes of 0157 were tested against *E. coli* chromosomal DNA of pools 1 to 19, 29 and 30 (Table 9). For *wbdN*, three pairs of primers, which bind to various regions of this gene, were tested and found to be specific for 0157 as there was no amplified DNA in any of those 21 pools of *E. coli* chromosomal DNA tested. Three pairs of primers for *wbdO* were tested, and they are all specific although primers # 1211/#1212 produced two or three bands of the wrong size from all pools. Three pairs of primers were tested for *wbdP* and they all were specific. Two pairs of primers were tested for *wbdR* and they were all specific. For *wzy*, three pairs of primers were tested and all were specific although primer pair #1203/#1204 produced one or three bands of the wrong size in each pool. For *wzx*, two pairs of primers were tested and both were specific although primer pair #1217/#1218 produced 2 bands of wrong size in 2 pools, and 1 band of wrong size in 7 pools. Bands of the wrong size found in amplified DNA are assumed to be due to chance hybridisation of genes widely present in *E. coli*. The primers, annealing temperatures and function information for each gene are in Table 9.

The 0157 assay was also performed using pools 37 to 42, including DNA from O antigen expressing *Yersinia pseudotuberculosis*, *Shigella boydii*, *Yersinia enterocolitica* 09, *Brucella abortus* and *Salmonella enterica* strains (Table 9A). None of the oligonucleotides derived from *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* or *wbdR* reacted specifically with these pools, except that primer pair #1203/#1204 produced two bands with *Y. enterocolitica* 09 and one of the bands is of the same size with that from the positive control. Primer pair #1203/#1204 binds to *wzy*. The predicted secondary

structures of Wzy proteins are generally similar, although there is very low similarity at amino acid or DNA level among the sequenced wzy genes. Thus, it is possible that *Y. enterocolitica* O9 has a wzy gene closely related to that of *E. coli* O157. It is also possible that this band is due to chance hybridization of another gene, as the other two wzy primer pairs (#1205/#1206 and #1207/#1208) did not produce any band with *Y. enterocolitica* O9. Notably, pool number 37 includes *S. enterica* Landau which has the same O antigen as *E. coli* O157, and pool 38 and 39 contain DNA of *B. abortus* and *Y. enterocolitica* O9 which cross react serologically with *E. coli* O157. This result indicates that these genes are highly O157 specific, although one primer pair may have cross reacted with *Y. enterocolitica* O9.

Each of the 16 pairs binding to *wbdN*, *wzx*, *wzy*, *wbdO*, *wbdP* and *wbdR* produces a band of predicted size with the pools containing O157 DNA (pools number 31 to 36). As pool 29 included DNA from all strains present in pools 31 to 36 other than O157 strain DNA (Table 6), we conclude that the 16 pairs of primers all give a positive PCR test with each of the five unrelated O157 strains.

Thus PCR using primers based on genes *wbdN*, *wzy*, *wbdO*, *wzx*, *wbdP* and *wbdR* is highly specific for *E. coli* O157, giving positive results with each of six unrelated O157 strains while only one primer pair gave a band of the expected size with one of three strains with O antigens known to cross-react serologically with *E. coli* O157.

TABLE 1

H7 strains used in this work in addition to the H
antigens type strains

5

| Name used in this study | Serotype | Original name | Source* |
|-------------------------------|----------|------------------|---------|
| M527 | O157:H7 | C664-1992 | a |
| M917 | O18ac:H7 | A57 | IMVS |
| M918 | O18ac:H7 | A62 | IMVS |
| M973 | O2:H7 | A1107 | CDC |
| M1004 | O157:H7 | EH7 | b |
| M1179 | O18ac:H7 | D-M3291/54 | IMVS |
| M1200 | O7:H7 | A64 | c |
| M1211 | O19ab:H7 | F8188-41 | IMVS |
| M1328 | O53:H7 | 14097 | IMVS |
| M1686 | O55:H7 | TB156 | d |

*

10

a. Statens Serum Institut, Copenhagen, Denmark.

b. Dr R. Brown of Royal Children's Hospital, Melbourne, Australia.

15

c. Max-Planck Institut fur molekulare Genetik, Berlin, Germany.

d. Dr P. Tarr of Children's Hospital and Medical Center, University of Washington, USA.

20

IMVS, Institute of Medical and veterinary Science, Adelaide, Australia.

CDC, Centers for Disease Control and prevention, Atlanta, USA.

Table 2
Oligonucleotides used to PCR amplify *fliC* genes
from different H type strains for sequencing

| H Type Strains | Annealing Temperature (°C) | Primers Used |
|----------------|----------------------------|--------------|
| 1 | 55 | #1575/#1576 |
| 2 | 55 | #1285/#1286 |
| 3 | 55 | #1285/#1286 |
| 4 | 50 | #1431/#1432 |
| 5 | 60 | #1285/#1286 |
| 6 | 55 | #1575/#1576 |
| 7 | 55 | #1575/#1576 |
| 8 | 55 | #1431/#1432 |
| 9 | 60 | #1575/#1576 |
| 10 | 55 | #1575/#1576 |
| 11 | 55 | #1285/#1286 |
| 12 | 60 | #1575/#1576 |
| 14 | 60 | #1575/#1576 |
| 15 | 60 | #1575/#1576 |
| 16 | 60 | #1575/#1576 |
| 17 | 60 | #1417/#1418 |
| 18 | 60 | #1575/#1576 |
| 19 | 60 | #1575/#1576 |
| 20 | 60 | #1575/#1576 |
| 21 | 55 | #1285/#1286 |
| 23 | 60 | #1575/#1576 |
| 24 | 60 | #1285/#1286 |
| 25 | 60 | #1417/#1418 |
| 26 | 60 | #1575/#1576 |
| 27 | 50 | #1431/#1432 |
| 28 | 60 | #1575/#1576 |
| 29 | 60 | #1285/#1286 |
| 30 | 60 | #1575/#1576 |
| 31 | 60 | #1575/#1576 |
| 32 | 60 | #1575/#1576 |
| 33 | 60 | #1285/#1286 |
| 34 | 55 | #1575/#1576 |
| 35 | 50 | #1431/#1432 |
| 37 | 60 | #1285/#1286 |
| 38 | 60 | #1285/#1286 |
| 39 | 55 | #1285/#1286 |
| 40 | 55 | #1285/#1286 |
| 41 | 60 | #1575/#1576 |
| 42 | 60 | #1285/#1286 |
| 43 | 60 | #1575/#1576 |
| 44 | 60 | #1285/#1286 |
| 45 | 60 | #1575/#1576 |
| 46 | 60 | #1575/#1576 |
| 47 | 55 | #1285/#1286 |
| 48 | 60 | #1575/#1576 |
| 49 | 60 | #1575/#1576 |
| 50 | 60 | #1285/#1286 |
| 51 | 60 | #1575/#1576 |
| 52 | 60 | #1575/#1576 |
| 54 | 50 | #1431/#1432 |
| 55 | 60 | #1285/#1286 |
| 56 | 60 | #1285/#1286 |

Table 3
Summary of the flagellin sequences obtained and specific H type
oligonucleotide primers

| H type strain(s) the sequenced gene(s) obtained from | H specificity coded by the gene(s) | H type strain from which the flagellin gene sequence was used for primer choice | Positions of primer 1 | Positions of primer 2 |
|--|------------------------------------|---|---------------------------|-----------------------------|
| 1 | 1 | 1 | 892-909 | 1172-1189 |
| 2 | 2 | 2 | 568-587 | 1039-1056 |
| 4,17,44 | 4 | 4 | 466-483 | 628-648 |
| 5 | 5 | 5 | 697-714 | 877-897 |
| 6 | 6 | 6 | 565-585 | 799-816 |
| 7 | 7 | 7 | 553-570 (primer #1806) | 1483-1500 (primer #1809) |
| 9 | 9 | 9 | 616-633 | 838-855 |
| 10(50)*** | 10 | 10 | 559-579 | 697-717 |
| 11 | 11 | 11 | 586-606* | 791-810* |
| 12 | 12 | 12 | 892-909 | 1172-1189 |
| 14 | 14 | 14 | 586-606 | 793-813 |
| 15 | 15 | 15 | 640-660 | 817-834 |
| 3 | 16 | 3 | 649-666 | 925-942 |
| 18 | 18 | 18 | 589-606 | 802-819 |
| 19 | 19 | 19 | 607-624 | 538-855 |
| 20 | 20 | 20 | 574-591 | 760-780 |
| 21,47 | 21 | 21 | 676-693** | 862-879** |
| 23 | 23 | 23 | 637-654 | 1336-1353 |
| 24 | 24 | 24 | 496-516 | 772-792 |
| 26 | 26 | 26 | 553-570 | 772-789 |
| 27 | 27 | 27 | 685-702 | 799-819 |
| 28 | 28 | 28 | 592-609 | 778-798 |
| 29 | 29 | 29 | 538-555 | 757-774 |
| 30 | 30 | 30 | 814-831 | 943-962 |
| 31 | 31 | 31 | 571-588 | 790-807 |
| 32 | 32 | 32 | 514-831 | 1057-1074 |
| 33 | 33 | 33 | 553-570 | 718-735 |
| 34 | 34 | 34 | 568-585 | 796-816 |
| 38,55 | 38 | 38 | 553-573 | 709-729 |
| 39 | 39 | 39 | 556-573 | 718-735 |
| 41 | 41 | 41 | 598-615 | 784-801 |
| 42 | 42 | 42 | 547-567 | 715-735 |
| 43 | 43 | 43 | 580-597 | 844-861 |
| 45 | 45 | 45 | 640-657 | 943-963 |
| 46 | 46 | 46 | 565-582 | 781-801 |
| 49 | 49 | 49 | 589-609 | 754-771 |
| 51 | 51 | 51 | 565-582 | 1042-1059 |
| 52 | 52 | 52 | 598-615 | 829-846 |
| 56 | 56 | 56 | 697-714 | 877-897 |
| 8 and 40 | | 8 | 562-579 | 1045-1062 |
| 25 | | 25 | 529-549 | 703-723 |
| 35 | | non-functional H11 gene | 769-789* | 1045-1065* |
| 37 | | 37 | 520-537 | 715-735 |
| 48 | | 48 | 568-585 | 835-852 |
| 54 | | non-functional H21 gene | 988-1008** | 1344-1364** |

* See section 13 for choice of primers for the flagellin gene of H11

** See section 13 for choice of primers for the flagellin gene of H21

*** See text

Table 3A
Cloning, expression and identification of flagellin genes

| H type strain from which the H antigen gene was amplified | Primers used for PCR amplification of the H antigen gene | Annealing temperature (oC) used for PCR amplification | Plasmid carrying the H antigen gene | Host strain used for expression | Anti-serum which reacts with an <i>E. Coli</i> <i>fliC</i> deletion strain carrying the plasmid | H antigen encoded by the cloned gene |
|---|--|---|-------------------------------------|---------------------------------|---|--------------------------------------|
| H1 | #1868 & #1870 | 55 | pPR1920 | M2126 | H1 | H1 |
| H2 | #1868 & #1870 | 55 | pPR1977 | P5560 | H2 | H2 |
| H3 | #1868 & #1870 | 55 | pPR1969 | P5560 | H16 | H16 |
| H4 | #1878 & #1885 | 65 | pPR1955 | P5560 | H4 | H4 |
| H5 | #1868 & #1870 | 60 | pPR1967 | M2126 | H5 | H5 |
| H6 | #1868 & #1870 | 55 | pPR1921 | P5560 | H6 | H6 |
| H7 | #1868 & #1870 | 55 | pPR1919 | P5560 | H7 | H7 |
| H9 | #1868 & #1870 | 55 | pPR1922 | P5560 | H9 | H9 |
| H10 | #1868 & #1870 | 55 | pPR1923 | P5560 | H10 | H10 |
| H11 | #1868 & #1870 | 55 | pPR1981 | M2126 | H11 | H11 |
| H12 | #1868 & #1870 | 60 | pPR1990 | M2126 | H12 | H12 |
| H14 | #1868 & #1870 | 55 | pPR1924 | P5560 | H14 | H14 |
| H15 | #1868 & #1870 | 55 | pPR1925 | P5560 | H15 | H15 |
| H17 | #1878 & #1885 | 65 | pPR1957 | P5560 | H4 | H4 |
| H18 | #1868 & #1870 | 55 | pPR1986 | M2126 | H18 | H18 |
| H19 | #1868 & #1870 | 55 | pPR1927 | P5560 | H19 | H19 |
| H20 | #1868 & #1870 | 55 | pPR1963 | M2126 | H20 | H20 |
| H21 | #1868 & #1870 | 55 | pPR1995 | M2126 | H21 | H21 |
| H23 | #1868 & #1869 | 55 | pPR1942 | P5560 | H23 | H23 |
| H24 | #1868 & #1870 | 55 | pPR1971 | M2126 | H24 | H24 |
| H26 | #1868 & #1870 | 65 | pPR1928 | P5560 | H26 | H26 |
| H27 | #1868 & #1870 | 55 | pPR1970 | M2126 | H27 | H27 |
| H28 | #1868 & #1870 | 60 | pPR1944 | P5560 | H28 | H28 |
| H29 | #1868 & #1870 | 55 | pPR1972 | M2126 | H29 | H29 |
| H30 | #1868 & #1871 | 55 | pPR1948 | P5560 | H30 | H30 |
| H31 | #1868 & #1870 | 65 | pPR1965 | M2126 | H31 | H31 |
| H32 | #1868 & #1871 | 55 | pPR1940 | P5560 | H32 | H32 |
| H33 | #1868 & #1871 | 55 | pPR1976 | M2126 | H33 | H33 |
| H34 | #1868 & #1870 | 65 | pPR1930 | P5560 | H34 | H34 |
| H38 | #1868 & #1870 | 48 | pPR1984 | M2126 | H38 | H38 |
| H39 | #1868 & #1870 | 48 | pPR1982 | M2126 | H39 | H39 |
| H41 | #1868 & #1870 | 65 | pPR1931 | P5560 | H41 | H41 |
| H42 | #1868 & #1870 | 50 | pPR1979 | M2126 | H42 | H42 |
| H43 | #1868 & #1870 | 65 | pPR1968 | M2126 | H43 | H43 |
| H45 | #1868 & #1870 | 60 | pPR1943 | P5560 | H45 | H45 |
| H46 | #1868 & #1870 | 60 | pPR1966 | M2126 | H46 | H46 |
| H49 | #1868 & #1870 | 60 | pPR1985 | M2126 | H49 | H49 |
| H51 | #1868 & #1870 | 65 | pPR1941 | P5560 | H51 | H51 |
| H52 | #1868 & #1870 | 65 | pPR1935 | P5560 | H52 | H52 |
| H56 | #1868 & #1870 | 50 | pPR1978 | M2126 | H56 | H56 |

Table 3B Oligonucleotide primers used for PCR amplification and cloning of H antigen genes

#1868 5'- cat gcc atg gca caa gtc att aat acc -3'
NcoI

#1869 5'- ata tgt cga ctt aac cct gca gca gag aca g -3'
SalI

#1870 5' - atg gat cct taa ccc tgc agc aga gac ag -3'
BamHI

#1871 5' - aac tgc agt taa ccc tgt agc aga gac ag -3'
PstI

#1872 5' - cgg gat ccc gca gac tgg ttc ttg ttg at - 3'
BamHI

#1878 5' - cgg gat cca ctt cta tcg agc gcc tct ct - 3'
BamHI

#1884 5' - gct cta gag cgc aga tca ttc agc agg cc -3'
XbaI

#1885 5' - gct cta gac atg ttg gac act tcg gtc gc - 3'
XbaI

- 75 -

TABLE 4

| Pool No. | Strains of which chromosomal DNA included in the pool | Source* |
|----------|---|-------------------|
| 1 | <i>E. coli</i> type strains for O serotypes 1, 2, 3, 4, 10, 16, 18 and 39 | IMVS ^a |
| 2 | <i>E. coli</i> type strains for O serotypes 40, 41, 48, 49, 71, 73, 88 and 100 | IMVS |
| 3 | <i>E. coli</i> type strains for O serotypes 102, 109, 119, 120, 121, 125, 126 and 137 | IMVS |
| 4 | <i>E. coli</i> type strains for O serotypes 138, 139, 149, 7, 5, 6, 11 and 12 | IMVS |
| 5 | <i>E. coli</i> type strains for O serotypes 13, 14, 15, 17, 19ab, 20, 21 and 22 | IMVS |
| 6 | <i>E. coli</i> type strains for O serotypes 23, 24, 25, 26, 27, 28, 29 and 30 | IMVS |
| 7 | <i>E. coli</i> type strains for O serotypes 32, 33, 34, 35, 36, 37, 38 and 42 | IMVS |
| 8 | <i>E. coli</i> type strains for O serotypes 43, 44, 45, 46, 50, 51, 52 and 53 | IMVS |
| 9 | <i>E. coli</i> type strains for O serotypes 54, 55, 56, 57, 58, 59, 60 and 61 | IMVS |
| 10 | <i>E. coli</i> type strains for O serotypes 62, 63, 64, 65, 66, 68, 69 and 70 | IMVS |
| 11 | <i>E. coli</i> type strains for O serotypes 74, 75, 76, 77, 78, 79, 80 and 81 | IMVS |
| 12 | <i>E. coli</i> type strains for O serotypes 82, 83, 84, 85, 86, 87, 89 and 90 | IMVS |
| 13 | <i>E. coli</i> type strains for O serotypes 91, 92, 95, 96, 97, 98, 99 and 101 | IMVS |
| 14 | <i>E. coli</i> type strains for O serotypes 103, 104, 105, 106, 107, 108 and 110 | IMVS |
| 15 | <i>E. coli</i> type strains for O serotypes 112, 162, 113, 114, 115, 116, 117 and 118 | IMVS |
| 16 | <i>E. coli</i> type strains for O serotypes 123, 165, 166, 167, 168, 169, 170 and 171 | See b |
| 17 | <i>E. coli</i> type strains for O serotypes 172, 173, 127, 128, 129, 130, 131 and 132 | See c |
| 18 | <i>E. coli</i> type strains for O serotypes 133, 134, 135, 136, 140, 141, 142 and 143 | IMVS |
| 19 | <i>E. coli</i> type strains for O serotypes 144, 145, 146, 147, 148, 150, 151 and 152 | IMVS |

*

a. Institute of Medical and Veterinary Science, Adelaide, Australia

b. 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark

c. 172 and 173 from Statens Serum Institut, Copenhagen, Denmark, the rest from IMVS

TABLE 5

| Pool No. | Strains of which chromosomal DNA included in the pool | Source* |
|----------|--|---------|
| 20 | <i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 157, 158, 159 and 160 | IMVS |
| 21 | <i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9 and 124 | IMVS |
| 22 | As pool #21, plus <i>E. coli</i> 0111 type strain Stoke W. | IMVS |
| 23 | As pool #21, plus <i>E. coli</i> 0111:H2 strain C1250-1991 | See d |
| 24 | As pool #21, plus <i>E. coli</i> 0111:H12 strain C156-1989 | See e |
| 25 | As pool #21, plus <i>S. enterica</i> serovar Adelaide | See f |
| 26 | <i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII | See g |
| 27 | <i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15 | See h |
| 28 | <i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z,15 and 52:d:e,n,x,z15 | IMVS |

*

- d. C1250-1991 from Statens Serum Institut, Copenhagen, Denmark
- e. C156-1989 from Statens Serum Institut, Copenhagen, Denmark
- f. *S. enterica* serovar Adelaide from IMVS
- g. Dr S Aleksic of Institute of Hygiene, Germany
- h. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada

TABLE 6

| Pool No. | Strains of which chromosomal DNA included in the pool | Source* |
|----------|---|----------------|
| 29 | <i>E. coli</i> type strains for O serotypes 153, 154, 155, 156, 158, 159 and 160 | IMVS |
| 30 | <i>E. coli</i> type strains for O serotypes 161, 163, 164, 8, 9, 111 and 124 | IMVS |
| 31 | As pool #29, plus <i>E. coli</i> O157 type strain A2 (O157:H19) | IMVS |
| 32 | As pool #29, plus <i>E. coli</i> O157:H16 strain C475-89 | See d |
| 33 | As pool #29, plus <i>E. coli</i> O157:H45 strain C727-89 | See d |
| 34 | As pool #29, plus <i>E. coli</i> O157:H2 strain C252-94 | See d |
| 35 | As pool #29, plus <i>E. coli</i> O157:H39 strain C258-94 | See d |
| 36 | As pool #29, plus <i>E. coli</i> O157:H26 | See e |
| 37 | As pool #29, plus <i>S. enterica</i> serovar Landau | See f |
| 38 | As pool #29, plus <i>Brucella abortus</i> | See g See h |
| 39 | As pool #29, plus <i>Y. enterocolitica</i> O9 | |
| 40 | <i>Y. pseudotuberculosis</i> strains of O groups IA, IIA, IIB, IIC, III, IVA, IVB, VA, VB, VI and VII | See i |
| 41 | <i>S. boydii</i> strains of serogroups 1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14 and 15 | See j |
| 42 | <i>S. enterica</i> strains of serovars (each representing a different O group) Typhi, Montevideo, Ferruch, Jangwani, Raus, Hvittingfoss, Waycross, Dan, Dugbe, Basel, 65:i:e,n,z15 and 52:d:e,n,x,z15 | IMVS |
| 43 | <i>E. coli</i> type strains for O serotypes 1,2,3,4,10,18 and 29 | IMVS |
| 44 | As pool #43, plus <i>E. coli</i> K-12 strains C600 and WG1 | IVMS See k |

*

- d. O157 strains from Statens Serum Institut, Copenhagen, Denmark
- e. O157:H26 from Dr R Brown of Royal Children's Hospital, Melbourne, Victoria
- f. *S. enterica* serovar Landau from Dr M Poppoff of Institut Pasteur, Paris, France
- g. *B. Abortus* from the culture collection of The University of Sydney, Sydney, Australia
- h. *Y. enterocolitica* O9 from Dr. K. Bettelheim of Victorian Infectious Diseases Reference Laboratory Victoria, Australia.
- i. Dr S Aleksic of Institute of Hygiene, Germany
- j. Dr J Lefebvre of Bacterial Identification Section, Laboratoire de Santé Publique du Québec, Canada
- k. Strains C600 and WG1 from Dr. B.J. Backmann of Department of Biology, Yale University, USA.

TABLE 7 PCR assay result using primers based on the E. coli serotype O16 (strain K-12) O antigen gene cluster sequence

| Gene | Function | Base positions of the gene | Forward primer (base positions) | Reverse primer (base positions) | Length of the PCR fragment | Number of pools (out of 21) giving band of correct size | Annealing temperature of the PCR |
|-----------------|------------------------------|----------------------------|---------------------------------|---------------------------------|----------------------------|---|----------------------------------|
| <i>rmlB</i> * | TDP-rhamnose pathway | 90-1175 | #1064(91-109) | #1065(1175-1157) | 1085bp | 17 | 60°C |
| <i>rmlD</i> * | TDP-rhamnose pathway | 1175-2074 | #1066(1175-1193) | #1067 (2075-2058) | 901bp | 13 | 60°C |
| <i>rmlA</i> * | TDP-rhamnose pathway | 2132-3013 | #1068(2131-2148) | #1069(3013-2995) | 883bp | 0 | 60°C |
| <i>rmlC</i> * | TDP-rhamnose pathway | 3013-3570 | #1070(3012-3029) | #1071(3570-3551) | 559bp | 0 | 60°C |
| <i>gjf</i> * | Galactofuranose pathway | 4822-5925 | #1074(4822-4840) | #1075(5925-5908) | 1104bp | 0 | 55°C |
| <i>wzx</i> * | Flippase | 3567-4814 | #1072(3567-3586) | #1073(4814-4797) | 1248bp | 0 | 55°C |
| <i>wzy</i> * | O polymerase | 5925-7091 | #1076(5925-5944) | #1077(7091-7074) | 1167bp | 0 | 60°C |
| <i>wbbI</i> * | Galactofuranosyl transferase | 7094-8086 | #1078 (7094-7111) | #1079(8086-8069) | 993bp | 0 | 50°C |
| <i>wbbJ</i> * | Acetyltransferase | 8067-8654 | #1080(8067-8084) | #1081(8654-8632) | 588bp | 0 | 60°C |
| <i>wbbK</i> ** | Glucosyl transferase | 5770-6888 | #1082(5770-5787) | #1083(6888-6871) | 1119bp | 0 | 55°C |
| <i>wbbL</i> *** | Rhamanosyltransferase | 679-1437 | #1084(679-697) | #1085(1473-1456) | 795bp | 0**** | 55°C |

* ** *** Base positions based on GenBank entry U09876, U03041 and L19537 respectively
 19 pools giving a band of wrong size

TABLE 8 PCR assay data using 0111 primers

| Gene | Base positions of the gene according to SEQ ID NO: 1 | Forward primer (base positions) | Reverse primer (base positions) | Length of the PCR fragment | Number of pools (out of 21) giving band of correct size | Annealing temperature of the PCR |
|-------------|--|---------------------------------|---------------------------------|----------------------------|---|----------------------------------|
| <i>wbdH</i> | 739-1932 | #866 (739-757) | #867(1941-1924) | 1203bp | 0 | 60°C |
| | | #976(925-942) | #978(1731-1714) | 807bp | 0 | 60°C |
| | | #976(925-942) | #979(1347-1330) | 423bp | 0 | 60°C |
| | | #977(1165-1182) | #978(1731-1714) | 567bp | 0 | 60°C |
| <i>wzx</i> | 8646-9911 | #969(8646-8663) | #970(9908-9891) | 1263bp | 0 | 50°C |
| | | #1060(8906-8923) | #1062(9468-9451) | 563bp | 0 | 60°C |
| | | #1061(9150-9167) | #1063 (9754-9737) | 605bp | 0 | 50°C |
| <i>wzy</i> | 9901-10953 | #900(9976-9996) | #901(10827-10807) | 852bp | 0 | 60°C |
| | | #980(10113-10130) | #983(10484-10467) | 372bp | 0* | 61°C |
| <i>wbdL</i> | 10931-11824 | #870(10931-10949) | #871(11824-11796) | 894bp | 7 | 60°C |
| <i>wbdM</i> | 11821-12945 | #868(11821-11844) | #869(12945-12924) | 1125bp | 0 | 60°C |
| | | #984(12042-12059) | #987(12447-12430) | 406bp | 0 | 60°C |
| | | #985(12258-12275) | #986(12698-12681) | 441bp | 0** | 65°C |

* Giving a band of wrong size in all pools

** One pool giving a band of wrong size

TABLE 8A PCR specificity test data using 0111 primers

| Gene | Base positions of the gene according to SEQ ID NO: 1 | Forward primer (base positions) | Reverse primer (base positions) | Length of the PCR fragment | Number of pools (pools no. 25-28) giving band of correct size | Annealing temperature of the PCR |
|-------------|--|---------------------------------|---------------------------------|----------------------------|---|----------------------------------|
| <i>wbdH</i> | 739-1932 | #866 (739-757) | #867(1941-1924) | 1203bp | 0* | 60°C |
| | | #976(925-942) | #978(1731-1714) | 807bp | 0 | 60°C |
| | | #976(925-942) | #979(1347-1330) | 423bp | 0 | 60°C |
| | | #977(1165-1182) | #978(1731-1714) | 567bp | 0 | 60°C |
| <i>wzX</i> | 8646-9911 | #969(8646-8663) | #970(9908-9891) | 1263bp | 0 | 55°C |
| | | #1060(8906-8923) | #1062(9468-9451) | 563bp | 0 | 60°C |
| | | #1061(9150-9167) | #1063 (9754-9737) | 605bp | 0* | 50°C |
| <i>wzy</i> | 9901-10953 | #900(9976-9996) | #901(10827-10807) | 852bp | 0 | 60°C |
| | | #980(10113-10130) | #983(10484-10467) | 372bp | 0** | 60°C |
| <i>wbdL</i> | 10931-11824 | #870(10931-10949) | #871(11824-11796) | 894bp | 0 | 60°C |
| <i>wbdM</i> | 11821-12945 | #868(11821-11844) | #869(12945-12924) | 1125bp | 0 | 60°C |
| | | #984(12042-12059) | #987(12447-12430) | 406bp | 0 | 60°C |
| | | #985(12258-12275) | #986(12698-12681) | 441bp | 0* | 65°C |

* 1 pool giving a band of wrong size

** 2 pools giving 3 bands of wrong sizes, 1 pool giving 2 bands of wrong sizes

TABLE 9 PCR results using primers based on the *E. coli* O157 sequence

| Gene | Function | Base position of the gene according to SEQ ID NO: 2 | Forward primer (base positions) | Reverse primer (base positions) | Length of the PCR fragment | Number of pools (out of 21) giving band of correct size | Annealing temperature of the PCR |
|-------------|--------------------|---|---------------------------------|---------------------------------|----------------------------|---|----------------------------------|
| <i>wbdN</i> | Sugar transferase | 79-861 | #1197(79-96) | #1198 (861-844) | 783 | 0 | 55°C |
| | | | #1199(184-201) | #1200(531-514) | 348 | 0 | 55°C |
| | | | #1201(310-327) | #1202(768-751) | 459 | 0 | 55°C |
| <i>wzy</i> | O antigen | 858-2042 | #1203(858-875) | #1204(2042-2025) | 1185 | 0* | 50°C |
| | | | #1205(1053-1070) | #1206(1619-1602) | 567 | 0 | 63°C |
| | | | #1207(1278-1295) | #1208(1913-1896) | 636 | 0 | 60°C |
| <i>wbdO</i> | Sugar transferase | 2011-2757 | #1209(2011-2028) | #1210(2757-2740) | 747 | 0 | 50°C |
| | | | #1211(2110-2127) | #1212(2493-2476) | 384 | 0** | 62°C |
| | | | #1213(2305-2322) | #1214(2682-2665) | 378 | 0 | 60°C |
| <i>wzx</i> | O antigen flippase | 2744-4135 | #1215(2744-2761) | #1216(4135-4118) | 1392 | 0 | 50°C |
| | | | #1217(2942-2959) | #1218(3628-3611) | 687 | 0*** | 63°C |
| <i>wbdP</i> | Sugar transferase | 5257-6471 | #1221(5257-5274) | #1222(6471-6454) | 1215 | 0 | 55°C |
| | | | #1223(5440-5457) | #1224(5973-5956) | 534 | 0 | 55°C |
| | | | #1225(5707-5724) | #1226(6231-6214) | 525 | 0 | 55°C |
| <i>wbdR</i> | N-acetyl | 13156-13821 | #1229(13261-13278) | #1230(13629-13612) | 369 | 0 | 55°C |
| | | | #1231(13384-13401) | #1232(13731-13714) | 348 | 0 | 60°C |

* 3 bands of wrong size in one pool, 1 band of wrong size in all other pools

** 3 bands of wrong sizes in 9 pools, 2 bands of wrong size in all other pools

*** 2 bands of wrong sizes in 2 pools, 1 band of wrong size in 7 pools

TABLE 9A PCR results using primers based on the *E. coli* O157 sequence

| Gene | Function | Base position of the gene according to SEQ ID NO: 2 | Forward primer (base positions) | Reverse primer (base positions) | Length of the PCR fragment | Number of pools (pools no. 37-42) giving band of correct size | Annealing temperature of the PCR |
|-------------|----------------------|---|---------------------------------|---------------------------------|----------------------------|---|----------------------------------|
| <i>wbdN</i> | Sugar transferase | 79-861 | #1197(79-96) | #1198 (861-844) | 783 | 0* | 55°C |
| | | | #1199(184-201) | #1200(531-514) | 348 | 0* | 55°C |
| | | | #1201(310-327) | #1202(768-751) | 459 | 0 | 61°C |
| <i>wzy</i> | O antigen | 858-2042 | #1203(858-875) | #1204(2042-2025) | 1185 | 1** | 50°C |
| | | | #1205(1053-1070) | #1206(1619-1602) | 567 | 0*** | 60°C |
| | | | #1207(1278-1295) | #1208(1913-1896) | 636 | 0 | 60°C |
| <i>wbdO</i> | Sugar transferase | 2011-2757 | #1209(2011-2028) | #1210(2757-2740) | 747 | 0 | 50°C |
| | | | #1211(2110-2127) | #1212(2493-2476) | 384 | 0**** | 61°C |
| | | | #1213(2305-2322) | #1214(2682-2665) | 378 | 0 | 60°C |
| <i>wzx</i> | O antigen flippase | 2744-4135 | #1215(2744-2761) | #1216(4135-4118) | 1392 | 0 | 50°C |
| | | | #1217(2942-2959) | #1218(3628-3611) | 687 | 0 | 63°C |
| <i>wbdP</i> | Sugar transferase | 5257-6471 | #1221(5257-5274) | #1222(6471-6454) | 1215 | 0 | 55°C |
| | | | #1223(5440-5457) | #1224(5973-5956) | 534 | 0* | 60°C |
| | | | #1225(5707-5724) | #1226(6231-6214) | 525 | 0 | 55°C |
| <i>wbdR</i> | N-acetyl transferase | 13156-13821 | #1229(13261-13278) | #1230(13629-13612) | 369 | 0 | 50°C |
| | | | #1231(13384-13401) | #1232(13731-13714) | 348 | 0 | 60°C |

* 1 band of wrong size in one pool

** pool #39 giving two bands, one band of correct size, the other band of wrong size in another pool.

*** 2 bands of wrong sizes in one pool

**** 3 bands of wrong sizes in 2 pools, 2 bands of wrong sizes in 2 other pools

CLAIMS:

1. A nucleic acid molecule encoding all or part of an *E. coli* flagellin protein, provided that the nucleic acid molecule does not encode a protein expressed by the *E. coli* H1, H7, H12 or H48 type strains.

2. A nucleic acid molecule according to claim 1 wherein the molecule is derived from a *fliC* gene.

3. A nucleic acid molecule including all or part of a sequence according to any one of SEQ ID NOs:1 to 68.

4. A nucleic acid molecule consisting of all or part of a sequence according to any one of SEQ ID NOs: 1 to 68.

5. A nucleic acid molecule according to any one of claims 1-4 wherein the molecule is from about 10 to 20 nucleotides in length.

6. A nucleic acid molecule according to claim 5 wherein the molecule is capable of hybridising to the central region of a flagellin gene from which the molecule is derived.

7. A nucleic acid molecule selected from the group of nucleic acid molecules shown in Table 3.

8. A method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one nucleic acid molecule derived from a flagellin gene, wherein the at least one nucleic acid molecule is specific for a particular flagellin gene associated with the H serotype, to any *E. coli* in the sample which contain the gene, and detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically

hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

5 9. A method according to claim 8 wherein the at least one nucleic acid molecule is according to any one of claims 1 to 7.

10 10. A method according to claim 8 wherein the specifically hybridised nucleic acid molecules are detected by Southern Blot analysis.

15 11. A method of detecting the presence of *E. coli* of a particular H serotype in a sample, the method comprising the step of specifically hybridising at least one pair of nucleic acid molecules to any *E. coli* in the sample which contains the flagellin gene for the particular H serotype, wherein at least one of the nucleic acid molecules is specific for the particular flagellin gene associated with the H serotype, and detecting any
20 specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the H serotype in the sample.

25 12. A method according to claim 11 wherein the at least one pair of nucleic acid molecules is according to any one of claims 1 to 7.

30 13. A method according to claim 11 wherein the specifically hybridised nucleic acid molecules are detected by the polymerase chain reaction.

35 14. A method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

(a) specifically hybridising at least one nucleic acid molecule derived from and specific for a gene

encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen, to any *E. coli* in the sample which contain the gene;

(b) specifically hybridising at least one nucleic acid molecule derived from and specific for a particular flagellin gene associated with that H serotype, to any *E. coli* in the sample which contain the gene; and

(c) detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the particular H serotype and O serotype of *E. coli* in the sample.

15. A method according to claim 14 wherein the at least one nucleic acid molecule of step (a) is selected from the group consisting of:

wbdH (nucleotide position 739 to 1932 of Figure 5),
wzx (nucleotide position 8646 to 9911 of Figure 5),
wzy (nucleotide position 9901 to 10953 of Figure 5),
wbdM (nucleotide position 11821 to 12945 of Figure 5),
wbdN (nucleotide position 79 to 861 of Figure 6),
wbdO (nucleotide position 2011 to 2757 of Figure 6),
wbdP (nucleotide position 5257 to 6471 of Figure 6),
wbdR (nucleotide position 13156 to 13821 of Figure 6),
wzx (nucleotide position 2744 to 4135 of Figure 6) and
wzy (nucleotide position 858 to 2042 of Figure 6).

16. A method according to claim 14 wherein the at least one nucleic acid molecule of step (a) is selected from the group of nucleic acid molecules shown in Tables 8, 8A, 9 and 9A.

17. A method according to claim 14 wherein the at least one nucleic acid molecule of step (b) is according to any one of claims 1 to 7.

18. A method according to claim 14 wherein the specifically hybridised nucleic acid molecules are detected by Southern Blot analysis.

5

19. A method for detecting the presence of a particular O serotype and H serotype of *E. coli* in a sample, the method comprising the following steps:

10 (a) specifically hybridising at least one pair of nucleic acid molecules derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen, to any *E.*
15 *coli* in the sample which contain the gene;

(b) specifically hybridising at least one pair of nucleic acid molecules derived from and specific for a particular flagellin gene associated with that H serotype, to any *E. coli* in the sample which contain the
20 gene; and

(c) detecting any specifically hybridised nucleic acid molecules, wherein the presence of specifically hybridised nucleic acid molecules identifies the presence of the particular H serotype and O serotype of *E. coli* in
25 the sample.

20. A method according to claim 19 wherein the at least one pair of nucleic acid molecules of step (a) is selected from the group consisting of:

30 *wbdH* (nucleotide position 739 to 1932 of Figure 5),
wzx (nucleotide position 8646 to 9911 of Figure 5),
wzy (nucleotide position 9901 to 10953 of Figure 5),
wbdM (nucleotide position 11821 to 12945 of Figure 5),
wbdN (nucleotide position 79 to 861 of Figure 6),
35 *wbdO* (nucleotide position 2011 to 2757 of Figure 6),
wbdP (nucleotide position 5257 to 6471 of Figure 6),
wbdR (nucleotide position 13156 to 13821 of Figure 6),

wzx (nucleotide position 2744 to 4135 of Figure 6) and
wzy (nucleotide position 858 to 2042 of Figure 6).

21. A method according to claim 19 wherein the at least
5 one pair of nucleic acid molecules of step (a) is
selected from the group of nucleic acid molecules shown
in Tables 8, 8A, 9 and 9A.

22. A method according to claim 19 wherein the at least
10 one nucleic acid molecule of step (b) is according to any
one of claims 1 to 7.

23. A method according to claim 19 wherein the
15 specifically hybridised nucleic acid molecules are
detected by the polymerase chain reaction.

24. A method for detecting the presence of a particular
O serotype and H serotype of *E. coli* in a sample, the
method comprising the following steps:

20 (a) specifically hybridising at least one nucleic
acid molecule derived from and specific for a gene
encoding a flagellin associated with a particular *E. coli*
H antigen serotype, to any *E. coli* carrying the gene and
present in the sample;

25 and

(b) detecting the at least one specifically
hybridised nucleic acid molecule, wherein the at least one
nucleic acid molecule is specific for the particular
combination of O and H antigen.

30 25. A method according to claim 24 wherein the at least
nucleic acid molecule is according to any one of SEQ ID
NOS: 9, 55, 57 to 65.

35 26. A method for testing a food derived sample for the
presence of one or more particular *E. coli* O antigens and
H antigens, wherein the particular *E. coli* O and H

antigens in the food derived sample are detected using the method of any one of claims 8, 11, 14 or 19.

27. A method for testing a faecal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens wherein the particular *E. coli* O and H antigens in the faecal derived sample are detected using the method of any one of claims 8, 11, 14 or 19.

28. A method for testing a patient or animal derived sample for the presence of one or more particular *E. coli* O antigens and H antigens wherein the particular *E. coli* O and H antigens in the patient or animal derived sample are detected using the method of any one of claims 8, 11, 14 or 19.

29. A kit for identifying the H serotype of *E. coli*, the kit comprising at least one nucleic acid molecule according to any one of claims 1 to 7.

30. A kit for identifying the H and O serotype of *E. coli*, the kit comprising:

(a) at least one nucleic acid molecule derived from and specific for an *E. coli* flagellin gene; and

(b) at least one nucleic acid molecule derived from and specific for a gene encoding a transferase or a gene encoding an enzyme for the transport or processing of a polysaccharide or oligosaccharide unit, the gene being involved in the synthesis of a particular *E. coli* O antigen.

31. A kit according to claim 30 wherein the at least one nucleic acid molecule of (a) is selected from the group consisting of:

wbdH (nucleotide position 739 to 1932 of Figure 5),
wzx (nucleotide position 8646 to 9911 of Figure 5),
wzy (nucleotide position 9901 to 10953 of Figure 5),

5 *wbdM* (nucleotide position 11821 to 12945 of Figure 5),
 wbdN (nucleotide position 79 to 861 of Figure 6),
 wbdO (nucleotide position 2011 to 2757 of Figure 6),
 wbdP (nucleotide position 5257 to 6471 of Figure 6),
 wbdR (nucleotide position 13156 to 13821 of Figure 6),
 wzx (nucleotide position 2744 to 4135 of Figure 6) and
 wzy (nucleotide position 858 to 2042 of Figure 6).

10 32. A kit according to claim 30 wherein the at least one
 nucleic acid molecule of (a) is selected from the group
 of nucleic acid molecules shown in Tables 8, 8A, 9 and
 9A.

15 33. A kit according to claim 30 wherein the at least one
 nucleic acid molecule of (b) is according to any one of
 claims 1 to 7.

1/96

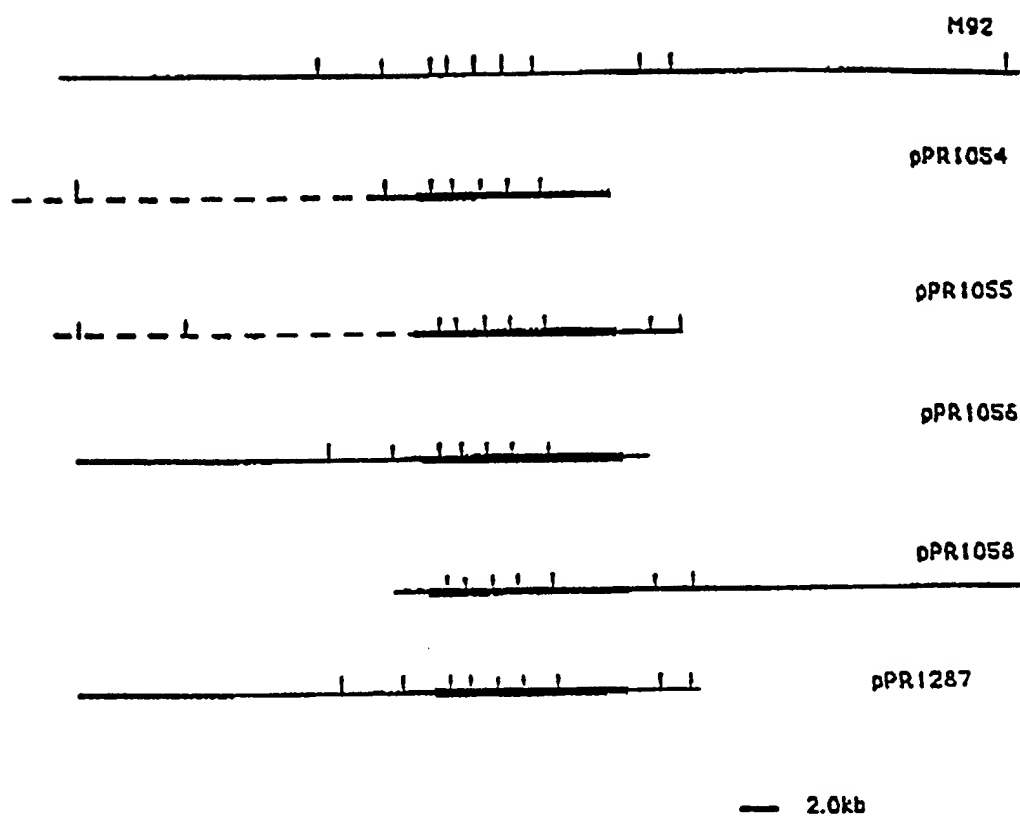


Figure 1

2/96

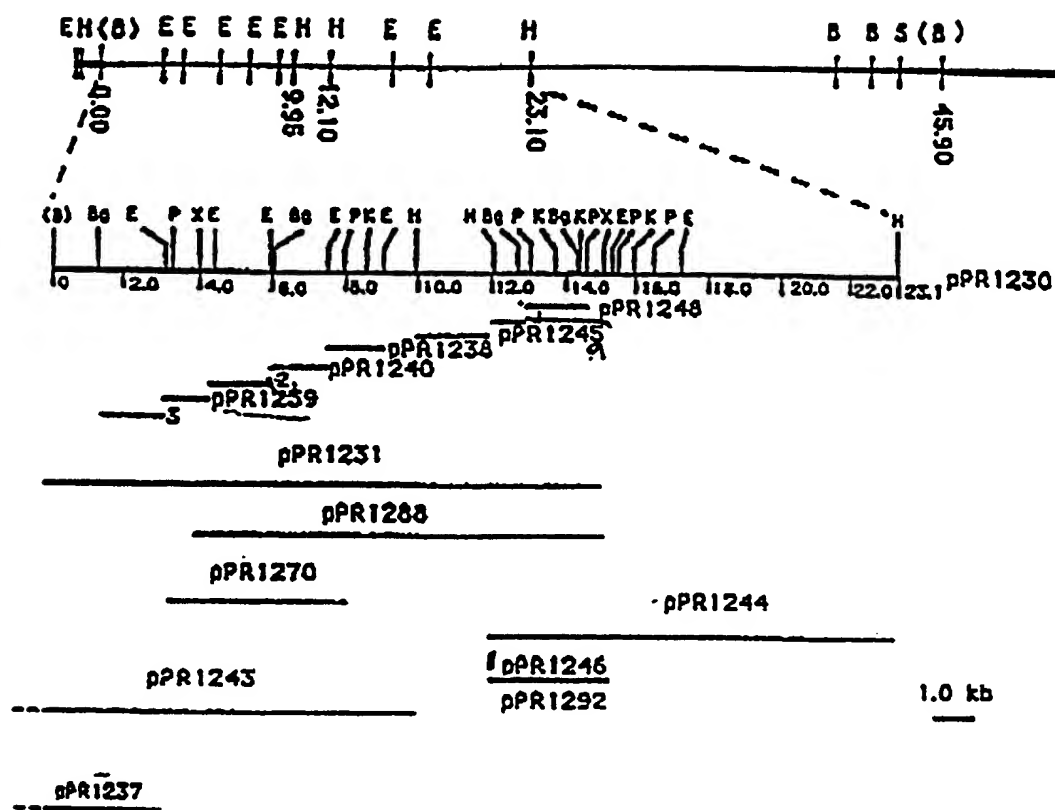


Figure 2

3/96

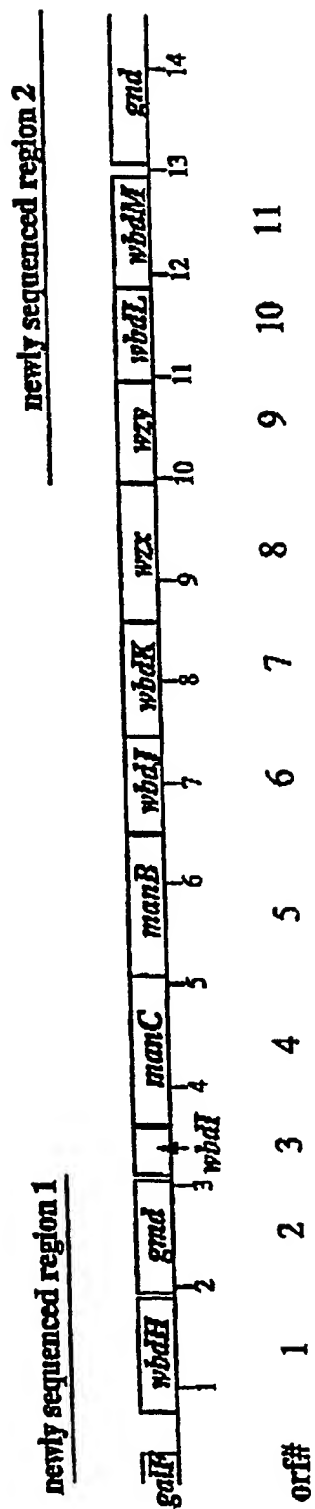


Figure 3

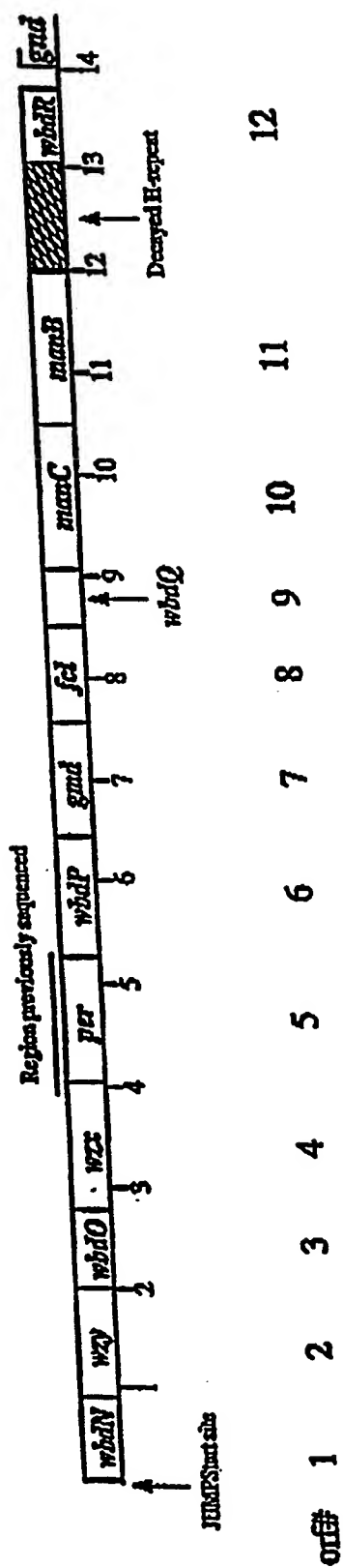


Figure 4

5/96

GATCTGATGGCCGTAGGGCGCTACGTGCTTTCTGCTGATATCTGGGCTGAGTTGGAAAAA 60

ACTGCTCCAGGTGCCTGGGGACGTATTCAACTGACTGATGCTATTGCAGAGTTGGCTAAA 120

AAACAGTCTGTTGATGCCATGCTGATGACCGGCGACAGCTACGACTGCGGTAAGAAGATG 180

GGCTATATGCAGGCATTCGTTAAGTATGGGCTGCGCAACCTTAAAGAAGGGGCGAAGTTC 240

CGTAAGAGCATCAAGAAGCTACTGAGTGAGTAGAGATTTACACGTCTTTGTGACGATAAG 300

CCAGAAAAAATAGCGGCAGTTAACATCCAGGCTTCTATGCTTTAAGCAATGGAATGTTAC 360

TGCCGTTTTTTATGAAAAATGACCAATAATAACAAGTTAACCTACCAAGTTTAATCTGCT 420

TTTTGTTGGATTTTTCTTGTCTTGGTCGCATTTGGTAAGACAATTAGCGTGAGTTTAA 480

GAGAGTTTTGCGGGATCTCGCGGAAGTCTCACATCTTTGGCATTTAGTTAGTGCAGTGG 540

TAGCTGTTAAGCCAGGGGCGGTAGCTTGCCTAATTAATTTTAACTATACATTTATTCT 600

TGCCGCTTATAGCAAATAAAGTCAATCGGATTAACTTCTTTTCCATTAGGTAAAAGAGT 660

GTTTGTAGTCGCTCAGGGAAATTGGTTTTGGTAGTAGTACTTTTCAAATTATCCATTTTC 720

Start of orf1

M L L C C I H I N V Y Y L L

CGATTTAGATGGCAGTTGATGTTACTATGCTGCATACATATCAATGTATATTATTTACTT 780

L E C D M K K I V I I G N V A S M M L R

TTAGAATGTGATATGAAAAAATAGTGATCATAGGCAATGTAGCGTCAATGATGTTAAGG 840

F R K E L I M N L V R Q G D N V Y C L A

TTCAGGAAAGAATTAATCATGAATTTAGTGAGGCAAGGTGATAATGTATATTGTCTAGCA 900

N D F S T E D L K V L S S W G V K G V K

AATGATTTTTTCCACTGAAGATCTTAAAGTACTTTTCGTATGGGGCGTTAAGGGGGTTAAA 960

F S L N S K G I N P F K D I I A V Y E L

TTCTCTCTTAAGTCAAAGGGTATTAATCCTTTTAAGGATATAATTGCTGTTTATGAACATA 1020

K K I L K D I S P D I V F S Y F V K P V

AAAAAATTCTTAAGGATATTTCCCAGATATTGTATTTTCATATTTTGTAAGCCAGTA 1080

I F G T I A S K L S K V P R I V G M I E

ATATTTGGAAGTATTGCTTCAAAGTTGTCAAAGTGCCAAGGATTGTTGGAATGATTGAA 1140

G L G N A F T Y Y K G K Q T T K T K M I

GGTCTAGGTAATGCCTTCACTTATTATAAGGGAAAGCAGACCACAAAACTAAATGATA 1200

K W I Q I L L Y K L A L P M L D D L I L

AAGTGGATACAAATCTTTTATATAAGTTAGCATTACCGATGCTTGATGATTTGATTCTA 1260

L N H D D K K D L I D Q Y N I K A K V T

TTAAATCATGATGATAAAAAAGATTTAATCGATCAGTATAATATTAAAGCTAAGGTAACA 1320

V L G G I G L D L N E F S Y K E P P K E

GTGTTAGGTGGGATTGGATTGGATCTTAATGAGTTTTTCATATAAAGAGCCACCGAAAGAG 1380

K I T F I F I A R L L R E K G I F E F I

AAAAATACCTTTATTTTATAGCAAGGTTATTAAGAGAGAAAGGGATATTTGAGTTTATT 1440

E A A K F V K T T Y P S S E F V I L G G

GAAGCCGCAAAGTTCGTTAAGACAACCTATCCAAGTCTGAATTTGTAATTTTAGGAGGT 1500

6/96

F E S N N P F S L Q K N E I E S L R K E 1560
 TTTGAGAGTAATAATCCTTTCTCATTACAAAAAATGAAATTGAATCGCTAAGAAAAGAA
 H D L I Y P G H V E N V Q D W L E K S S 1620
 CATGATCTTATTTATCCTGGTCATGTGGAAAAATGTTCAAGATTGGTTAGAGAAAAGTTCT
 V F V L P T S Y R E G V P R V I Q E A M 1680
 GTTTTTGTTTTACCTACATCATATCGAGAAGGCGTACCAAGGGTGATCCAAGAAGCTATG
 A I G R P V I T T N V P G C R D I I N D 1740
 GCTATTGGTAGACCTGTAATAACAATAATGTACCTGGGTGTAGGGATATAATAAATGAT
 G V N G F L I P P F E I N L L A E K M K 1800
 GGGGTCAATGGCTTTTTGATACCTCCATTTGAAATTAATTTACTGGCAGAAAAAATGAAA
 Y F I E N K D K V L E M G L A G R K F A 1860
 TATTTTATTGAGAATAAAGATAAAGTACTCGAAATGGGGCTTGCTGGAAGGAAGTTTGCA
 E K N F D A F E K N N R L A S I I K S N 1920
 GAAAAAACTTTGATGCTTTTGAAAAAATAATAGACTAGCATCAATAATAAATCAAAT

 End of orf1
 N D F *
 AATGATTTTTTGACTTGAGCAGAAATTATTTATATTTCAATCTGAAAAATAAAGGCTGTTA 1980

 Start of orf2
 M N K V A L I T G I T G Q D G S Y L A 2040
 TTATGAATAAAGTGGCATTAAATTACTGGTATCACTGGGCAAGATGGCTCCTATTTGGCAG
 E L L L E K G Y E V H G I K R R A S S F 2100
 AATTATTGTTAGAAAAAGGTTATGAAGTTCATGGTATTAAACGCCGTGCATCTTCATTTA
 N T E R V D H I Y Q D S H L A N P K L F 2160
 ATACTGAGCGAGTGGATCACATCTATCAGGATTCACATTTAGCTAATCCTAAACTTTTTC
 L H Y G D L T D T S N L T R I L K E V Q 2220
 TACACTATGGCGATTTGACAGATACTTCCAATCTGACCCGTATTTTAAAAGAAGTTCAAC
 P D E V Y N L G A M S H V A V S F E S P 2280
 CAGATGAAGTTTACAATTTGGGGGCGATGAGCCATGTAGCGGTATCATTTGAGTCACCAG
 E Y T A D V D A I G T L R L L E A I R I 2340
 AATACACTGCTGATGTTGATGCGATAGGAACATTGCGTCTTCTTGAAGCTATCAGGATAT
 L G L E K K T K F Y Q A S T S E L Y G L 2400
 TGGGGCTGGAAAAAAGACAAAATTTTATCAGGCTTCAACTTCAGAGCTTTATGGTTTGG
 V Q E I P Q K E T T P F Y P R S P Y A V 2460
 TTCAAGAAATTCCACAAAAAGAGACTACGCCATTTTATCCACGTTGCGCTTATGCTGTTG
 A K L Y A Y W I T V N Y R E S Y G M F A 2520
 CAAAATTATATGCCTATTGGATCACTGTTAATTATCGTGAGTCTTATGGTATGTTTGCCT
 C N G I L F N H E S P R R G E T F V T R 2580
 GCAATGGTATTCTCTTTAACCACGAATCACCTCGCCGTGGCGAGACCTTTGTTACTCGTA
 K I T R G I A N I A Q G L D K C L Y L G 2640
 AAATAACACGCGGGATAGCAAATATTGCTCAAGGTCTTGATAAATGCTTATACTTGGGAA
 N M D S L R D W G H A K D Y V K M Q W M 2700
 ATATGGATTCTCTGCGTGATTGGGGACATGCTAAGGATTATGTCAAAATGCAATGGATGA

7/96

M L Q Q E T P E D F V I A T G I Q Y S V
 TGCTGCAGCAAGAACTCCAGAAGATTTTGTAAATTGCTACAGGAATTCAATATTCTGTCC 2760
 R E F V T M A A E Q V G I E L A F E G E
 GTGAGTTTGTGCACAATGGCGGCAGAGCAAGTAGGCATAGAGTTAGCATTTGAAGGTGAGG 2820
 G V N E K G V V V S V N G T D A K A V N
 GAGTAAATGAAAAAGGTGTTGTTGTTTCGGTCAATGGCACTGATGCTAAAGCTGTAAACC 2880
 P G D V I I S V D P R Y F R P A E V E T
 CGGGCGATGTAATTATATCTGTAGATCCAAGGTATTTTAGGCCTGCAGAAGTTGAAACCT 2940
 L L G D P T N A H K K L G W S P E I T L
 TGCTTGGCGATCCTACTAATGCGCATAAAAAATTAGGATGGAGCCCTGAAATTACATTGC 3000
 R E M V K E M V S S D L A I A K K N V L
 GTGAAATGGTAAAAGAAATGGTTTCCAGCGATTAGCAATAGCGAAAAAGAACCTTTGC 3060

End of orf2

L K A N N I A T N I P Q E *
 TGAAGCTAATAACATTGCCACTAATATTCCGCAAGAA TAAAAAGATAATACATTAAAT 3120

Start of orf3

AATTAAAAATCGTCTAGATTATTAGTACCATTATTTTTTTTTTGGGTGACTAATGTTTA 3180
 I T S D K F R E I I K L V P L V S I D L
 TTACATCAGATAAATTTAGAGAAATTATCAAGTTAGTTCCATTAGTATCAATTGATCTGC 3240
 L I E N E N G E Y L F G L R N N R P A K
 TAATTGAAAACGAGAAATGGTGAATATTTATTTGGTCTTAGGAATAATCGACCGGCAAAA 3300
 N Y F F V P G G R I R K N E S I K N A F
 ATTATTTTTTTTTTGTCCAGGTGGTAGGATTGCAAAAATGAATCTATTAAAAATGCTTTTA 3360
 K R I S S M E L G K E Y G I S G S V F N
 AAAGAATATCATCTATGGAATTAGGTAAAGAGTATGGTATTTTCAGGAAGTCTTTTTTAATG 3420
 G V W E H F Y D D G F F S E G E A T H Y
 GTGTATGCGAACAATTTCTATGATGATGGTTTTTTTTTCTGAAGGCGAGGCAACACATTATA 3480
 I V L C Y T L K V L K S E L N L P D D Q
 TAGTGTCTTTGTTACACACTGAAGTTCTTAAAGTGAATTGAATCTCCCAAGATGATCAAC 3540
 H R E Y L W L T K H Q I N A K Q D V H N
 ATCGTGAATACCTTTGGCTAACTAAACACCAATAAATGCTAAACAAGATGTTTCATAACT 3600

End of orf3

Start of orf4

Y S K N Y F L * M
 ATTCAAAAAATTATTTTTTGTAAATTTTTTATTAATAATTAATATGCGAGAGAAATTGTATGT 3660
 S Q C L Y P V I I A G G T G S R L W P L
 ETCATCTCTTTTACCCTGTAAATTATGCGCGAGGAACCGGAAGCGTCTATGCGCGTTGT 3720
 S R V L Y P K Q F L N L V G D S T M L Q
 ETCGAGTATTATACCTTAACAATTTTAAATTTAGTTGGGGATTCTACAATGTTGCAAA 3780
 T T I T R L D G I E C E N P I V I C N E
 CAACAATTACCGCTTTGGATGGCATEGAATGCGAAATTECAATTGTTATCTGCAATGAAG 3840
 D H R F I V A E Q L R Q I G K L T K N I
 ATCAACGATTTATTGTAGCAGAGCAATTACGACAGATTGGTAAGCTAACCAAGAATATTA 3900
 I L E P K G R N T A P A I A L A A F I A
 TACTTGAGCCGAAAGGCCGTAACTGCACTGCACTAGCTTTAGCTGCTTTTATCGCTC 3960

8/96

Q K N N P N D D P L L L V L A A D H S I
 AGAAGAATAATCCTAATGACGACCCTTTATTATTAGTACTTGGGGCAGACCCTCTATAA 4020
 N N E K A F R E S I I K A M P Y A T S G
 ATAATGAAAAAGCATTTCGAGAGTCAATAATAAAGCTATGCGGTATGCAACTTCTGGGA 4080
 K L V T F G I I P D T A N T G Y G Y I K
 AGTTAGTAACATTTGGAATTATTCCGGACACGGGCAATACTGGTTATGGATATATTAGA 4140
 R S S S A D P N K E F P A Y N V A E F V
 GAAGTTCTTCAGCTGATCCTAATAAGAATTCCCAAGCATATAATGTTGGGGAGTTTGTAG 4200
 E K P D V K T A Q E Y I S S G N Y Y W N
 AAAAACAGATGTTAAAAACAGCACAGGAATATATTTTCGAGTGGGAATTATTACTGGAATA 4260
 S G M F L F R A S K Y L D E L R K F R P
 GCGGAATGTTTTTATTTCGGGGCAGTAATATCTTGATGAACCTACCGAATTTAGACCAG 4320
 D I Y H S C E C A T A T A N I D M D F V
 ATATTTTATCATAGCTGTGAATGTGCAACCGCTACAGCAAAATATAGATATGGACTTTGTCC 4380
 R I N E A E F I N C P E E S I D Y A V M
 GAATTAACGAGGCTGAGTTTATTAAATTGTCTCTGAAGAGTCTATCGATTATGCTGTGATGG 4440
 E K T K D A V V L P I D I G W N D V G S
 AAAAACAAAGACGCTGTAGTTCCTTCGGATAGATATTGGCTGGAATGACCTGGCTTCTT 4500
 W S S L W D I S Q K D C H G N V C H G D
 GGTCAATCACTTTGGGATATAAGCCAAAGGATTGGCATGGTAATGTGTGCCATGGGGATG 4560
 V L N H D G E N S F I Y S E S S L V A T
 TGCTCAATCATGATGGAGAAAATAGTTTATTATTTACTCTGAGTCAAGTCTGGTTGGGACAG 4620
 V G V S N L V I V Q T K D A V L V A D R
 TCGGAGTAAGTAATTTAGTAATTTGTCAAACCAAGGATGCTGTAAGTGGTTGGGGACCGTG 4680
 D K V Q N V K N I V D D L K K R K R A E
 ATAAAGTCCAAATGTTAAAAACATAGTTGACGATCTAAAAAGAGAAAAGCTGCTGAAT 4740
 Y Y M H R A V F R P W G K F D A I D Q G
 ACTACATGCATGCTGCAAGTTTTCGGCCCTTGGGGTAATTCGATGCAATAGACCAAGGCG 4800
 D R Y R V K K I I V K P G E G L D L R M
 ATAGATATAGAGTAATAAAAAATAATAGTTAAACCAAGGAGAAAGGTTAGATTTAAGGATGC 4860
 H H H R A E H W I V V S G T A K V S L G
 ATCATCATAGGGCAGAGCATTGGATTGTTGTATCCGGTACTGCTAAAGTTTCACTAGCTA 4920
 S E V K L L V S N E S I Y I P Q G A K Y
 GTGAAGTTAACTATTAGTTTCTAATGAGTCTATATATATCCCTCAGGGAGCAAAATATA 4980
 S L E N P G V I P L H L I E V S S G D Y
 GTCTTGAGAAATCCAGGCGTAATACCTTTGCATCTAATTCGAAGTAAGTTCTGCTGATTACC 5040
 L E S D D I V R F T D R Y N S K Q F L K
 TTGAATCAGATGATATAGTCCGTTTTACTGACAGATATAACAGTAACAAATTCCTAAGC 5100

End of orf4 Start of orf5

M N K I T C F K A Y D I R G R L
 R D *
 GAGATTGATAAATATGAATAAAATAACTTGCTTCAAAAGCATATGATATACGTGGGCGCTCT 5160

9/96

G A E L N D E I A Y R I G R A Y G E F F 5220
 TGGTGGCTGAATTGAATGATGAATAGCATATAGAATTGGGTGCGGCTTATGGTGAGTPTPTT
 K P Q T V V V G G D A R L T S E S L K K 5280
 TAAACCTCAAACTGTAGTTGTGGGAGGAGATGCTCGCTTAACAAGTGAGAGTTTAAAGAA
 S L S N G L C D A G V N V L D L G M C G 5340
 ATCACTCTCAAAATGGGCTATGTGATGCAGGCGTAAATGTCTTAGATCTTGGAAATGTGTGG
 T E E I Y F S T W Y L G I D G G I E V T 5400
 TACTGAAGAGATATATTTTCCACTTGGTATTTAGGAATTGATGCTGGAATCGAGGTAAC
 A S H N P I D Y N G M K L V T K G A R P 5460
 TGCAGGCCATAATCCAAATGATTATAATGGAATGAAATTAGTAACCAAGGTGCTCGACC
 I S S D T G L K D I Q Q L V E S N N F E 5520
 AATCAGCAGTGACACAGGTCTCAAGATATACAACAATTACTAGAGAGTAATAATTTTGA
 E L N L E K K G N I T K Y S T R D A Y I 5580
 AGAGCTCAACCTAGAAAAAAGGGAATATTACCAATATTCCACCCGAGATGCCCTACAT
 N H L M G Y A N L Q K I K K I K I V V N 5640
 AAATCATTGTGATGGGCTATGCTAATCTGCAAAAAATAAAAAATCAAAATAGTTGTGA
 S G N G A A G P V I D A I E E C F L R N 5700
 TTCTGGGAATGGTGCAGCTGGTCTCTTATTGATGCTATTGAGGAATGCTTTTACGGAA
 N I P I Q F V K I N N T P D G N F P H G 5760
 CAATATTCCGATTTCAGTTTGTAAAAATAAATAATACACCCGATGGTAATTTTCCACATGG
 I P N P L L P E C R E D T S S A V I R H 5820
 TATCCCTAATCCATTACTACCTGAGTGCAGAGAAGATACCAGCAGTGCCTTTATAAGACA
 S A D F G I A F D G D F D R C F F F D E 5880
 TACTGCTGATTTTGGTATTGCAATTTGATGCTGATTTTGTATAGGTCTTTTTTCTTTGATGA
 N G Q F I E G Y Y I V G L L A E V F L G 5940
 AAATGGACAATTTATTGAAGGATACTACATTCTTGGTTTATTAGCGGAAGTTTTTTTAGG
 K Y P N A K I I H D P R L I W N T I D I 6000
 GAAATATCCAAACGCAAAATCATTCATGATCCTCGCCTTATATGGAATACTATTGATAT
 V E S H G G I P I M T K T G H A Y I K Q 6060
 CGTAGAAAGTCATGCTGGTATACCTATAATGACTAAACCGCTCATGCTTACATTAAGCA
 R M R E E D A V Y G G E M S A H H Y F K 6120
 AAGAATGCCGTGAAGAGGATGCCGTATATGCCGCGCAATGAGTGGCGCATCATTATTTTAA
 D F A Y C D S G M I P W I L I C E L L S 6180
 AGATTTTGCATACTGCGATAGTGGAAATGATTCCTTGGATTTTAAATTTGTGAACTTTTGAG
 L T N K K L G E L V C G C I N D W P A S 6240
 TCTGACAAATAAAAAATTAGGTGAACGCTTGTGGTTCTATAAACGACTGGCCCGGCAAG
 G E I N C T L D N P Q N E I D K L F N R 6300
 TGGAGAAATAAACTGTACACTAGACAATCCGCAAAATGAAATAGATAAATTATTTAATCG
 Y K D S A L A V D Y T D G L T M E F S D 6360
 TTACAAAGATAGTGCCTTAGCTGTGTGATTACACTGATGGATTAACTATGGAGTTCTCTGA
 W R F N V R C S N T E P V V R L N V E S 6420
 TTGGCGTTTTTAATGTAGATGCTCAATACAGAACCTGTAGTACGATTGAATGTAGAATC
 R N N A I L M Q E K T E E I L N F I S K 6480
 TAGGAATAATGCTATTCTTATGCGAGGAAAAACAGAAGAAATTCTGAATTTTATATCAAA

10/96

| End of orf5 | Start of orf6 | |
|--|---------------|------|
| * ATAAATTTGCACCTGAGTTTCATAATGGGAACAAGAAATATATGAAAGTACTTCTGACTGG | M K V L L T G | 6540 |
| S T G M V G K N I L E H D; S A S K Y N I | | |
| CTCAACTGGCATGGTTGGTAAGAATATATTAGAGCATGATAGTGCAGTAAGTAAATATAATAT | | 6600 |
| L T P T S S D L N L L D K N E I E K F M | | |
| ACTTACTCCAACCAGCTCTGATTTGAATTTATTAGATAAAATGAAATAGAAAAATTCAT | | 6660 |
| L I N M P D C I I H A A G L V G G I H A | | |
| GCTTATCAACATGCCAGACTGTATTATACATGCAGCGGGATTAGTTGGAGGCATTTCATGC | | 6720 |
| N I S R P F D F L E K N L Q M G L N L V | | |
| AAATATAAGCAGGCGCTTTGATTTTCTGGAAGAAATTTGCAGATGGGTTTAAATTTAGT | | 6780 |
| S V A K K L G I K K V L N L G S S C M Y | | |
| TTCCGTCGCAAAAAAAGTAGGTATCAAGAAAGTGTAACTTGGGTAGTTCATGCATGTA | | 6840 |
| P K N F E E A I P E K A L L T G E L E E | | |
| CCCCAAAAAAGCTTTGAAGAGGCTATTTCCTGAGAAAGCTCTGTAACTGGTGAAGCTAGAAGA | | 6900 |
| T N E G Y A I A K I A V A K A C E Y I S | | |
| AACTATGAGGGATATGCTATTTCGGAAGAAATTCCTGTAGCAAAAGCATGCCAATATATATC | | 6960 |
| R E N S N Y F Y K T I I P C N L Y G K Y | | |
| AAGAGAAAACTCTAATTTATTTTATAAACAATTTATCCCATGTAAATTTATATCGGAAATA | | 7020 |
| D K F D D N S S H M I P A V I K K I H H | | |
| TGATAAATTTGATGATAACTCGTCAATATGATTCCGGGAGTTATAAAAAAATCCATCA | | 7080 |
| A K I N N V P E I E I W G D G N S R R E | | |
| TGCGAAAAATTAATAATCTCCCAAGAGATCGAAATTTGGGGGGATGGTAATTCGCGCCGCTGA | | 7140 |
| F M Y A E D L A D L I F Y V I P K I E F | | |
| GTTTATGTATGCAGAAAGATTTAGCTGATCTTATTTTTTATGTTATTCCTAAATAGAAAT | | 7200 |
| M P N M V N A G L G Y D Y S I N D Y Y K | | |
| CATGCCTAATATGGTAATGCTGCTTTAGCTTACGATTATTCAATTAATGACTATTATAA | | 7260 |
| I I A E E I G Y T G S F S H D L T K P T | | |
| GATAATTCGAGAAAGAAATTTGGTTATACTGGGAGTTTTTCTCATGATTTAACAAAACCAAC | | 7320 |
| G M K R K L V D I S L L N K I G W S S H | | |
| AGGAATGAAACGGAAGCTAGTAGATATTTCATTGCTTAATAAAATTTGCTTGGTCAAGTCA | | 7380 |
| F E L R D G I R K T Y N Y Y L E N Q N K | | |
| CTTTCAACTCAGAGATGGCATEAGAAAGACCTATAATTTATTACTTGGAGAAATCAAAATAA | | 7440 |

Start of orf7, End of orf6

| | |
|---|------|
| M I T Y P L A S N T W D E Y E Y A A I Q | |
| * ATGATTACATAACCACTTGGCTAGTAATACTTGGGGATGAATATGAGTATGCAGCAATACAG | 7500 |
| S V I D S K M F T M G K K V E L Y E K N | |
| TCAGTAATTTGACTCAAAAATGTTTACCATGGGTAAAAAGGTTGACTTATATGAGAAAAAT | 7560 |
| F A D L F G S K Y A V M V S S G S T A N | |
| TTTGCTGATTTGTTTTCGTAGCAAAATATGCCGTAATGGTTAGCTCTCGTTCTACAGCTAAT | 7620 |

11/96

L L M I A A L F F T N K P K L K R G D E
 CTGTTAATGATTGCTGCECTTTTCTTCACTAATAAACCAAACTTAAAGAGGTTGATGAA 7680
 I I V P A V S W S T T Y Y P L Q Q Y G L
 ATAATAGTACCTGCACTGTCATGGTCTACGACATATTACCCCTCTGCAACAGTATGGCTTA 7740
 K V K F V D I N K E T L N I D I D S L K
 AAGGTGAAGTTTGTGATATCAATAAAGAACTTTAAATATTGATATCGATAGTTTGAAA 7800
 N A I S D K T K A I L T V N L L G N P N
 AATGCTATTTCAGATAAAACAAAGCAATATTGACAGTAATTTATTAGGTAATCTCTAAT 7860
 D F A K I N E I I N N R D I I L L E D N
 GATTTTGCAAAAATAAATGAGATAATAAATAATAGGGATATTATCTTACTAGAAGATAAC 7920
 C E S M G A V F Q N K Q A G T F G V M G
 TGTGAGTGGATGGGCGCGGTCTTTCAAAATAAGCAGGCAGGCACATTCCGGAGTTATGGCT 7980
 T F S S F Y S H H I A T M E G G C V V T
 ACCTTTAGTTCTTTTCTTCTCTCATATAGCTACAAATGGAAGGGGGCTGCGTAGTTACT 8040
 D D E E L Y H V L L C L R A H G W T R N
 GATGATCAAGAGCTGTATCATGTATTGTTGCTGCTTGGAGCTCATGCTTGGACAAGAAAT 8100
 L P K E N M V T G T K S D D I F E E S F
 TTACCAAAAGAGAAATATGCTTACAGGCACTAAGAGTGAATGATATTTTCGAAGAGTCTGTT 8160
 K F V L P G Y N V R P L E M S G A I G I
 AAGTTTGTTTTACCAGGATACAATGTTTGGCCCACTTGAATGAGTGGTGTATTGGGATA 8220
 E Q L K K L P G F I S T R R S N A Q Y F
 GAGCAACTTAAAAAGTTACCAGGTTTTATATCCACCAGACGTTCCATGACCAATATTTT 8280
 V D K F K D H P F L D I Q K E V G E S S
 GTAGATAAATTTAAGATCATCCATTCCTTTCATATACAAAAGAGTTGGTGAAGTAGG 8340
 W F G F S F V I K E G A A I E R K S L V
 TGGTTTGGTTTCTTCTTCTTATAAAGGAGGGAGCTGCTATTGAGAGGAAGAGTTTAGTA 8400
 N N L I S A G I E C R P I V T G N F L K
 AATAATCTGATCTCAGCAGGCATTGAATGCGGACCAATGCTTACTGGGAATTTTCTCAAA 8460
 N E R V L S Y F D Y S V H D T V A N A E
 AATGAACGTTGTTTGGAGTTATTTTGAATTACTCTGTACATGATACGGTAGCAATGCGGAA 8520
 Y I D K N G F F V G N H Q I P L F N E I
 TATATAGATAAGAAATGGTTTCTTCTCGGAAACCACCAGATACCTTTGTTTAAATGAATA 8580
 D Y L R K V L K *
 End of orf7
 GATTATCTACGAAAAGTATTAAAAAATACTAACGAGGCACTCTATTTCGAATAGAGTGGCT 8640
 Start of orf8
 M V L T V K K I L A F G Y S K V L P
 TTAAGATGGTATTTAACAGTGAATAAAATTTTAGCGTTTGGCTATTCTAAAGTACTACCAC 8700
 P V I E Q F V N P I C I F I I T P L I L
 CGGTTATTGAACAGTTTGTCAATCCAATTTGCATCTTCATATACACCACTAATACTCA 8760
 N H L G K Q S Y G N W I L L I T I V S F
 ACCACCTGGGTAAAGCAAGCTATGGTAATTGGATTTTATTAATTACTATTGATCTTTT 8820

12/96

S Q L I C G G C S A W I A K I I A E Q R
 CTCAGTTAATATGTCGGAGGATGTTCCGCGATGGATTGCAAAATCATTGCGAGAACAGAGAA 8880
 I L S D L S K K N A L R Q I S Y N F S I
 TTCTTAGTGATTATATCAAAAAAATGCTTTACGTCAAATTTCTATATAATTTTCAATTG 8940
 V I I A F A V L I S F L I L S I C F F D
 TTATTATCCGATTTCGCGGTATTGATTCTCTTTCTTATATTAAGTATTTGTTTCTTCGATG 9000
 V A R N N S S F L F A I I I C G F F Q E
 TTCCGAGGAATAATTCTTCATTCTTATTCGCGATTATTATTTGTCGGTPTTTTTCAGGAAG 9060
 V D N L F S G A L K G F E K F N V S C F
 TTGATAATTTATTTTAGTGCTGCGCTAAAGGTTTTTGAAAAATTTAATGATCATGTTTTT 9120
 F E V I T R V L W A S I V I Y G I Y G N
 TTGAAGTAATTACAGAGTCTCTCGGCTTCTATAGTAATATATGCCATTTCAGGAATG 9180
 A L L Y F T C L A F T I K G M L K Y I L
 CACTCTTATATTTTACATGTTTAGCCTTTACCATTAAAGGTATGCTAAATATATTTCTTG 9240
 V C L N I T G C F I N P N F N R V G I V
 TATGCTCTGAATATTACCGGTTGTTTCATCAATCTTAATTTTAATAGAGTTGGGATTGTTA 9300
 N L L N E S K W M F L Q L T G G V S L S
 ATTTGTTAAATGAGTCAAAATCGATGPTTCTTCAATTAAGTGGTGGCTCTCACTTAGTT 9360
 L F D R L V I P L I L S V S K L A S Y V
 TGTTTGATAGGCTCGTAATACCATTGATTTTATCTGTCTAGTAAAGTGGCTTCTTATGTEE 9420
 P C L Q L A Q L M F T L S A S A N Q I L
 CTTGCECTTCAACTAGCTCAATTTGATGTTCACTCTTTCTGCGCTCTGCAATCAAAATATTAC 9480
 L P M F A R M K A S N T F P S N C F F K
 TACCAATGTTTTCGTAGAAATGAAAGCATCTAACACATTTCCCTCTAATTTGTTTTTTTAAA 9540
 I L L V S L I S V L P C L A L F F F G R
 TTCTGCTTGTATCACTAATTTCTGTTTTGCECTTGTCTTGGCTTATTCTTTTTTTTGGCTG 9600
 D I L S I W I N P T F A T E N Y K L M Q
 ATATATTATCAATATGGAATAAACCTACATTTGCAACTGAAAATTATAAATTAATGCAAA 9660
 I L A I S Y I L L S M M T S F H F L L L
 TTTTAGCTATAAGTTACATTTTATTGTCAATGATGACATCTTTTCAATTTCTTTGTTATTAG 9720
 G I G K S K L V A N L N L V A G L A L A
 GAATGCTAAATCTAAGCTTGTGCAAAATTTAAATCTGCTTGCAGGCTTCGCACTTGCTG 9780
 A S T L I A A H Y G L Y A I S M V K I I
 CTTCAACGTTAATGCGAGCTCAATTATGGCTTTATGCAATATCTATGCTAAAAATAATAT 9840
 Y P A F Q F Y Y L Y V A F V Y F N R A K
 ATCCGGCTTTTCAATTTTATTACCTTTATGTAGCTTTTGTCTATTTTAAATAGACGAAAA 9900

Start of orf9, End of orf8

M S I D L L F S I T E I A I V F S C T I
 N V Y *
 ATGCTCTATTGATTACTTTTTTCAATTAAGTGAATGCAATTTGTTTTTTCTTGCCTATT 9960
 Y I F T Q C L L M R R I Y L D K S I L I
 TACATATTTTACTCAATGTTTGTAAATGCGGAGGATCTATTTAGATAAAAGTATTTTAATT 10020
 L L C L L F F L V I I Q L P E L N V N G
 CTTTTATGCTTGCTCTTTTTTTTTAGTAATCATTCAACTTCCTGAGCTTAATGTAAACGGT 10080

13/96

L V D S L K L S L P L L M V F I A F Q K
 TTGGTCGATTCTTTAAAGTTATCACTGCCTTTATTGATGGTCTTTATCGCTTTTCAAAAA 10140
 P K L C L W V I I A L L F L N S A F N F
 CCGAAATTATGCTTGTGGGTATTATTGCATTGTTGTTTTTGAACCTCTGCATTTAATTTT 10200
 L Y L K T F D K F S S F P F T F F I L L
 TTATATTTAAAGACATTCGATAAGTTTAGCTCATTTCCTTTTACTTTTTTTATATTGCTG 10260
 F Y L F R L G I G N L P V Y K N K K F Y
 TTTTACTTGTTTAGATTGGGAATTGGTAATTTACCGGTTTATAAAAATAAAAAATTTTAC 10320
 A L I F L F I L I D I M Q S L L I N Y R
 GCGTTGATTTTCTCTTTATATTAATAGACATAATGCAGTCATTGTTAATAAATTATAGG 10380
 G Q I L Y S V I C I L I L V F K V N L R
 GGGCAGATTTTATATTCCGTAATTTGCATCCTGATACTTGTGTTTAAAGTTAATTTAAGA 10440
 K K I P Y F F L M L P V L Y V I I M A Y
 AAAAGATTCCATACTTTTTTTTAAATGCTGCCAGTTTATATGTAATTATTATGGCTTAT 10500
 I G F N Y F N K G V T F F E P T A S N I
 ATTGGTTTTAATTATTTCAATAAAGGCGTAACTTTTTTTGAACCTACAGCAAGTAATATT 10560
 E R T G M I Y Y L V S Q L G D Y I F H G
 GAACGTACGGGGATGATATATTATTTGGTTTCACAGCTTGGTGATTATATATTCCATGGT 10620
 M G T L N F L N N G G Q Y K T L Y G L P
 ATGGGGACATTAAATTTCTTAAATAACGGCGGACAATATAAGACGTTATATGGACTTCCA 10680
 S L I P N D P H D F L L R F F I S I G V
 TCATTAATTCCTAATGACCCTCATGATTTTTTTATTACGGTCTTTTATAAGTATTGGTGTG 10740
 I G A L V Y H S I F F V F F R R I S F L
 ATAGGAGCATTGGTTTATCATCTATATTTTTTTGTTTTTTTTTAGGAGAATATCTTTCTTA 10800
 L Y E R N A P F I V V S C L L L L Q V V
 TTATATGAGAGAAATGCTCCTTTCATTGTTGTAAGTTGTTTGTACTGTTACAAGTTGTG 10860
 L I Y T L N P F D A F N R L I C G L T V
 TTAATTTATACATTAAACCCTTTTGATGCTTTTAATCGATTGATTTGCGGGCTTACAGTT 10920

Start of orf10

End of orf9

G V V Y G F A K I R *
 M D L Q K L D K Y T C N G N L D A
 GGAGTTGTTTATGGATTGCAAAAATTAGATAAGTATACCTGTAATGGAAATTTAGACGC 10980
 P L V S I I I A T Y N S E L D I A K C L
 TCCACTTGTTCATAATCATTCGCAACTTATAATTCTGAACTTGATATAGCTAAGTGTTC 11040
 Q S V T N Q S Y K N I E I I I M D G G S
 GCAATCGGTAACATAATCAATCTTATAAGAATATTGAAATCATAATAATGGATGGAGGATC 11100
 S D K T L D I A K S F K D D R I K I V S
 TTCTGATAAAACGCTTGATATTGCAAAATCGTTTAAAGACGACCGAATAAAAAATAGTTTC 11160
 E K D R G I Y D A W N K A V D L S I G D
 AGAGAAAGATCGTGGAATTTATGATGCCCTGGAATAAAGCAGTTGATTTATCCATTGGTGA 11220
 W V A F I G S D D V Y Y H T D A I A S L
 TTGGGTAGCATTTATTGGTTCAGATGATGTTTACTATCATACAGATGCAATTGCTTCATT 11280
 M K G V M V S N G A P V V Y G R T A H E
 GATGAAGGGGGTTATGGTATCTAATGGCGCCCTGTGGTTTATGGGAGGACAGCGCACGA 11340

14/96

G P D R N I S G F S G S E W Y N L T G F
AGGTCCCGATAGGAACATATCTGGATTTTCAGGCAGTGAATGGTACAACCTAACAGGATT 11400

K F N Y Y K C N L P L P I M S A I Y S R
TAAGTTTAATTATTACAAATGTAATTTACCATTGCCCATTATGAGC\$CAATATATTCTCG 11460

D F F R N E R F D I K L K I V A D A D W
TGATTTCTTCAGAAACGAACGTTTGTATATTAAATTAAAAATTGTTGCTGACGCTGATTG 11520

F L R C F I K W S K E K S P Y F I N D T
GTTTCTGAGATGTTTCATCAAATGGAGTAAAGAGAAGTCACCTTATTTTATTAATGACAC 11580

T P I V R M G Y G G V S T D I S S Q V K
GACCCCTATTGTTAGAATGGGATATGGTGGGGTTTCGACTGATATTTCTTCTCAAGTTAA 11640

T T L E S F I V R K K N N I S C L N I Q
AACTACGCTAGAAAGTTTCATTGTACGCAAAAAGAATAATATATCCTGTTTAAACATACA 11700

L I L R Y A K I L V M V A I K N I F G N
GCTGATTCTTAGATATGCTAAATTTCTGGTGATGGTAGCGATCAAAAATATTTTGGCAA 11760

N V Y K L M H N G Y H S L K K I K N K I
TAATGTTTATAAATTAATGCATAACGGGTATCATTCCTTAAAGAAAATCAAGAATAAAAT 11820

Start of orf11, End of orf10

M K I V Y I I T G L T C G G A E H L M T
*
ATGAAGATTGTTTATATAATAACCGGGCTTACTTGTGGTGGAGCCGAACACCTTATGACG 11880

Q L A D Q M F I R G H D V N I I C L T G
CAGTTAGCAGACCAAATGTTTATACGCGGGCATGATGTTAATATTATTTGTCTAACTGGT 11940

I S E V K P T Q N I N I H Y V N M D K N
ATATCTGAGGTAAAGCCAACACAAAATATTAATATTCATTATGTTAATATGGATAAAAAT 12000

F R S F F R A L F Q V K K I I V A L K P
TTTAGAAGCTTTTTTAGAGCTTTATTTCAAGTAAAAAAATAATTGTCGCCTTAAAGCCA 12060

D I I H S H M F H A N I F S R F I R M L
GATATAATACATAGTCATATGTTTCATGCTAATATTTTGTAGTCGTTTATTAGGATGCTG 12120

I P A V P L I C T A H N K N E G G N A R
ATTCCAGCGGTGCCCCCTGATATGTACCGCACACAACAAAATGAAGGTGGCAATGCAAGG 12180

M F C Y R L S D F L A S I T T N V S K E
ATGTTTTGTTATCGACTGAGTGATTTTTTAGCTTCTATTACTACAAATGTAAGTAAAGAG 12240

A V Q E F I A R K A T P K N K I V E I P
GCTGTTCAAGAGTTTATAGCAAGAAAGGCTACACCTAAAAATAAAATAGTAGAGATTCCG 12300

N F I N T N K F D F D I N V R K K T R D
AATTTTATTAATAACAAATAAATTTGATTTTGATATTAATGTCAGAAAGAAAACGCGAGAT 12360

A F N L K D S T A V L L A V G R L V E A
GCTTTTAATTTGAAAGACAGTACAGCAGTACTGCTCGCAGTAGGAAGACTTGTTGAAGCA 12420

K D Y P N L L N A I N H L I L S K T S N
AAAGACTATCCGAACCTATTAAATGCAATAAATCATTTGATTCTTTCAAAAACATCAAAT 12480

C N D F I L L I A G D G A L R N K L L D
TGTAATGATTTTATTTTGTCTATTGCTGGCGATGGCGCATTAAGAAATAAATTATTGGAT 12540

L V C Q L N L V D K V F F L G Q R S D I
TTGGTTTGTCAATTGAATCTTGTGGATAAAGTTTCTTCTTGGGGCAAAGAAGTGATATT 12600

15/96

K E L M C A A D L F V L S S E W E G F G 12660
 AAAGAATTAATGTGTGCTGCAGATCTTTTGTGTTTGTGAGTTCTGAGTGGGAAGGTTTGGT
 L V V A E A M A C E R P V V A T D S G G 12720
 CTCGTTGTTGCAGAAGCTATGGCGTGTGAACGTCCCGTTGTTGCTACCGATTCTGGTGA
 V K E V V G P H N D V I P V S N H I L L 12780
 GTTAAAGAAGTCGTTGGACCTCATAATGATGTTATCCCTGTCAGTAATCATATTCTGTTG
 A E K I A E T L K I D D N A R K I I G M 12840
 GCAGAGAAAAATCGCTGAGACACTTAAAATAGATGATAACGCAAGAAAAATAATAGGTATG
 K N R E Y I V S N F S I K T I V S E W E 12900
 AAAAATAGAGAATATATTGTTTCCAATTTTCAATTAAAACGATAGTGAGTGAGTGGGAG

End of orf11

R L Y F K Y S K R N N I I D *
 CGCTTATATTTTAAATATTCCAAGCGTAATAATATAATTGAT TGAAAAATATAAGTTTGTA 12960
 CTCTGGATGCAATAGTTTCTCTATGCTGTTTTTTTTACTGGCTCCGTATTTTTTACTTATAG 13020
 CTGGATTTTGTTATATATCAGTATTAATCTGTCTCAACTTCATCTAGACTACATTCAAGC 13080

Start of gnd

CGCGCATGCGTCGCGCGGTGACTACACCTGACAGGAGTATGTA ATGTCCAAGCAACAGAT 13140
 M S K Q Q I
 G V V G M A V M G R N L A L N I E S R G 13200
 CGGCGTCGTCGGTATGGCAGTGATGGGGCGCAACCTGGCGCTCAACATCGAAAGCCGCGG
 Y T V S I F N R S R E K T E E V V A E N 13260
 TTATACCGTCTCCATCTTCAACCGCTCCCGCGAGAAAACCTGAAGAAGTTGTTGCCGAGAA
 P D K K L V P Y Y T V K E F V E S L E T 13320
 CCCGGATAAGAACTGGTTCCTTATTACACGGTGAAAGAGTTTCGTGAGTCTCTTGAAAC
 P R R I L L M V K A G A G T D A A I D S 13380
 CCCACGTCGTATCCTGTTAATGGTAAAAGCAGGGGCGGGAACCTGATGCTGCTATCGATT
 L K P Y L D K G D I I I D G G N T F F Q 13440
 CCTGAAGCCGTATCTGGATAAAGGCGACATCATTATTGATGGTGGCAACACCTTCTTCCA
 D T I R R N R E L S A E G F N F I G T G 13500
 GGACACTATCCGTCGTAACCGTGAACCTGTCCGCGGAAGGCTTTAACTTCATCGGTACCGG
 V S G G E E G A L K G P S I M P G G Q K 13560
 CGTGTCCGGCGGTGAAGAGGGCGCCCTGAAAGGCCCATCTATCATGCCAGGTGGCCAGAA
 E A Y E L V A P I L T K I A A V A E D G 13620
 AGAAGCGTATGAGCTGGTTGCGCCTATCCTGACCAAGATTGCTGCGGTTGCTGAAGATGG
 E P C I T Y I G A D G A G H Y V K M V H 13680
 CGAACCATGTATAACTTACATCGGTGCTGACGGTGCGGGTCACCTACGTGAAGATGGTGCA
 N G I E Y G D M Q L I A E A Y S L L K G 13740
 CAACGGTATCGAATATGGCGATATGCAGCTGATTGCTGAAGCCTATTCTCTGCTTAAAGG
 G L N L S N E E L A T T F T E W N E G E 13800
 CGGCCTTAATCTGTCTAACGAAGAGCTGGCAACCACTTTTACCGAGTGGAATGAAGGCGA
 L S S Y L I D I T K D I F T K K D E E G 13860
 GCTAAGTAGCTACCTGATTGACATCACCAAAGACATCTTCACCAAAAAAGATGAAGAGGG

16/96

K Y L V D V I L D E A A N K G T G K W T
TAAATACCTGGTTGATGTGATCCTGGACGAAGCTGCGAACAAAGGCACCGGTAAATGGAC 13920

S Q S S L D L G E P L S L I T E S V F A
CAGCCAGAGCTCTCTGGATCTGGGTGAACCGCTGTCGCTGATCACCGAATCCGTATTTCGC 13980

R Y I S S L K D Q R I A A S K V L S G P
TCGCTACATCTCTTCTCTGAAAGACCAGCGCATTGCGGCATCTAAAGTGCTGTCTGGTCC 14040

Q A K L A G D K A E F V E K V R R A L Y
GCAGGCTAAACTGGCTGGTGATAAAGCAGAGTTCGTTGAGAAAGTCCGTCGCGCGCTGTA 14100

L G K I V S Y A Q G F S Q L R A A S D E
CCTGGGTAAATCGTCTCTTATGCCCAAGGCTTCTCTCAACTGCGTGCCGCGTCTGACGA 14160

Y N W D L N Y G E I A K I F R A G C I I
ATACAACTGGGATCTGAACTACGGCGAAATCGCGAAGATCTTCCGCGCGGGCTGCATCAT 14220

R A Q F L Q K I T D A Y A E N K G I A N
TCGTGCGCAGTTCTTGCAGAAAATTACTGACGCGTATGCTGAAAACAAAGGCATTGCTAA 14280

L L L A P Y F K N I A D E Y Q Q A L R D
CCTGTTGCTGGCTCCGTACTTCAAAAATATCGCTGATGAATATCAGCAAGCGCTGCGTGA 14340

V V A Y A V Q N G I P V P T F S A A V A
TGTAGTGGCTTATGCTGTGCAGAACGGTATTCCGGTACCGACCTTCTCTGCAGCGGTAGC 14400

Y Y D S Y R S A V L P A N L I Q A Q R D
CTACTACGACAGCTACCGTTCTGCGGTACTGCCGGCTAATCTGATTTCAGGCACAGCGTGA 14460

Y F G A H T Y K R T D K E G V F H T G
TTACTTCGGTGCGCACACGTATAAACGCACTGATAAAGAAGGTGTGTTCCACACCG 14516

17/96

GTAACCAAGGGCGGTACGTGCATAAAATTTTAATGCTTATCAAAACTATTAGCATTAAAAA 60

Start of orf1

M N K E T V S I I M P V Y N
TATATAAGAAATTCTCAAATGAAGAAAGAAACCGTTTCAATAATTATGCCCGTTTACAAT 120

G A K T I I S S V E S I I H Q S Y Q D F
GGGGCCAAAACATAATCTCATCAGTAGAATCAATTATACATCAATCTTATCAAGATTTT 180

V L Y I I D D C S T D D T F S L I N S R
GTTTTGTATATCATTGACGATTGTAGCACCGATGATACATTTTCATTAATCAACAGTCGA 240

Y K N N Q K I R I L R N K T N L G V A E
TACAAAACAATCAGAAAATAAGAATATTGCGTAACAAGACAAATTTAGGTGTTGCAGAA 300

S R N Y G I E M A T G K Y I S F C D A D
AGTCGAAATTATGGAATAGAAATGGCCACGGGGAAATATATTTCTTTTGTGATGCGGAT 360

D L W H E K K L E R Q I E V L N N E C V
GATTTGTGGCAGCAGAAAAAATTAGAGCGTCAAATCGAAGTGTTAAATAATGAATGTGTA 420

D V V C S N Y Y V I D N N R N I V G E V
GATGTGGTATGTTCTAATTATTATGTTATAGATAACAATAGAAATATTGTTGGCGAAGTT 480

N A P H V I N Y R K M L M K N Y I G N L
AATGCTCCTCATGTGATAAATTATAGAAAAATGCTCATGAAAACTACATAGGGAATTTG 540

T G I Y N A N K L G K F Y Q K K I G H E
ACAGGAATCTATAATGCCAACAAATTGGGTAAGTTTATCAAAAAAGATTGGTGCACGAG 600

D Y L M W L E I I N K T N G A I C I Q D
GATTATTTGATGTGGCTGGAAATAATTAATAAAACAAATGGTGCTATTTGTATTCAAGAT 660

N L A Y Y M R S N N S L S G N K I K A A
AATCTGGCGTATTACATGCGTTCAAATAATTCATCTCGGGTAATAAAATTAAAGCTGCA 720

K W T W S I Y R E H L H L S F P K T L Y
AAATGGACATGGAGTATATATAGAGAACATTTACATTTGTCCTTTCCAAAAACATTATAT 780

Y F L L Y A S N G V M K K I T H S L L R
TATTTTTTATTATATGCTTCAAATGGAGTCATGAAAAAATAACACATTCATTAAGG 840

Start of orf2, End of orf1

R K E T K K *
V K S A A K L I F L F L F T
AGAAAGGAGACTAAAAAGTGAAGTCAGCGGCTAAGTTGATTTTTTTATTCCTATTTACAC 900

L Y S L Q L Y G V I I D D R I T N F D T
TTTATAGTCTCCAGTTGTATGGGGTTATCATAGATGATCGTATAACAAATTTTGATACAA 960

K V L T S I I I I F Q I F F V L L F Y L
AGGTATTAAGTAGTATTATAATTATATTTTTCAGATTTTTTTTGTATTATTATTTATCTAA 1020

T I I N E R K Q Q K K F I V N W E L K L
CGATTATAAATGAAAGAAAACAGCAGAAAAAATTATCGTGAACCTGGGAGCTAAAGTTAA 1080

I L V F L F V T I E I A A V V L F L K E
TACTCGTTTTTCCTTTTTGTGACTATAGAAATTGCTGCTGTAGTTTATTCTTAAAGAAG 1140

G I P I F D D D P G G A K L R I A E G N
GTATTCCTATATTTGATGATGATCCAGGGGGGGCTAAACTTAGAATAGCTGAAGGTAATG 1200

18/96

G L Y I R Y I K Y F G N I V V F A L I I
 GACTTTACATTAGATATATTAAGTATTTTGGTAATATAGTTGTGTTTGCATTAATTATTC 1260
 L Y D E H K F K Q R T I I F V Y F T T I
 TTTATGATGAGCATAAAATTCAAACAGAGGACCATCATATTTGTATATTTTACAACGATTG 1320
 A L F G Y R S E L V L L I L Q Y I L I T
 CTTTATTTGGTTATCGTTCTGAATTGGTGTGCTCATTCTTCAATATATATTGATTACCA 1380
 N I L S K D N R N P K I K R I I G Y F L
 ATATCCTGTCAAAGGATAACCGTAATCCTAAAATAAAAAGAATAATAGGGTATTTTTTAT 1440
 L V G V V C S L F Y L S L G Q D G E Q N
 TGGTAGGGGTTGTATGCTCGTTGTTTTATCTAAGTTTAGGACAAGACGGAGAACAAAATG 1500
 D S Y N N M L R I I N R L T I E Q V E G
 ACTCATATAATAATATGTTAAGGATAATTAATAGGTTAACAATAGAGCAAGTTGAAGGTG 1560
 V P Y V V S E S I K N D F F P T P E L E
 TTCCATATGTTGTTTCTGAATCTATTAAGAACGATTTCTTTCCGACACCAGAGTTAGAAA 1620
 K E L K A I I N R I Q G I K H Q D L F Y
 AGGAATTAAGCAATAATAAATAGAAATACAGGGAATAAAGCATCAAGACTTATTTTATG 1680
 G E R L H K Q V F G D M G A N F L S V T
 GAGAACGGTTACATAAACAAGTATTTGGAGACATGGGAGCAAATTTTTTATCAGTTACTA 1740
 T Y G A E L L V F F G F L C V F I I P L
 CGTATGGAGCAGAACTGTAGTTTTTTTGGTTTTCTCTGTGTATTTCATTATCCCTTTAG 1800
 G I Y I P F Y L L K R M K K T H S S I N
 GGATATATATACCTTTTTTATCTTTTAAAGAGAATGAAAAAACCCATAGCTCGATAAATT 1860
 C A F Y S Y I I M I L L Q Y L V A G N A
 GCGCATTCTATTTCATATATCATTATGATTTTATTGCAATACTTAGTGGCTGGGAATGCAT 1920
 S A F F F G P F L S V L I M C T P L I L
 CGGCCTTCTTTTTTGGTCCTTTTCTCTCCGTATTGATAATGTGTACTCCTCTGATCTTAT 1980

Start of orf3

M K I S V I T V T Y
 L H D T L K R L S R N E N I S Y N C D L
 TGCATGATACGTTAAAGAGATTATCACGAAATGAAAATATCAGTTATAACTGTGACTTAT 2040

End of orf2

N N A E G L E K T L S S L S I L K I K P
 *
 AATAATGCTGAAGGGTTAGAAAAAACTTTAAGTAGTTTATCAATTTTAAAAATAAAACCT 2100
 F E I I I V D G G S T D G T N R V I S R
 TTTGAGATTATTATAGTTGATGGCGGCTCTACAGATGGAACGAATCGTGTCTATTAGTAGA 2160
 F T S M N I T H V Y E K D E G I Y D A M
 TTTACTAGTATGAATATTACACATGTTTATGAAAAAGATGAAGGGATATATGATGCGATG 2220
 N K G R M L A K G D L I H Y L N A G D S
 AATAAGGGCCGAATGTTGGCCAAAGGCGACTTAATACATTATTTAAACGCCGCGGATAGC 2280
 V I G D I Y K N I K E P C L I K V G L F
 GTAATTGGAGATATATATAAAAAATATCAAAGAGCCATGTTTGATTAAAGTTGGCCTTTTC 2340
 E N D K L L G F S S I T H S N T G Y C H
 GAAAATGATAAACTTCTGGGATTTTCTCTATAACCCATTCAAATACAGGGTATTGTCAT 2400

19/96

Q G V I F P K N H S E Y D L R Y K I C A 2460
 CAAGGGGTGATTTTCCCAAAGAATCATTTCAGAAATATGATCTAAGGTATAAAATATGTGCT
 D Y K L I Q E V F P E G L R S L S L I T 2520
 GATTATAAGCTTATTCAAGAGGTGTTTCCTGAAGGGTTAAGATCTCTATCTTTGATTACT
 S G Y V K Y D M G G V S S K K R I L R D 2580
 TCGGGTTATGTAAAATATGATATGGGGGGAGTATCTTCAAAAAAAGAATTTTAAGAGAT
 K E L A K I M F E K N K K N L I K F I P 2640
 AAAGAGCTTGCCAAAATTATGTTTGAAAAAATAAAAAAACCTTATTAAGTTTATTCCA
 I S I I K I L F P E R L R R V L R K M Q 2700
 ATTTCAATAATCAAAATTTTATTCCCTGAACGTTTAAGAAGAGTATTGCGGAAAATGCAA

 Y I C L T L F F M K N S S P Y D N E *
 TATATTTGTCTAACTTTATTCTTCATGAAGAATAGTTCCACCATATGATAATGAATAAAAT 2760
 K K I L K F C T L K K Y D T S S A L G R 2820
 CAAAAAATACTTAAATTTTGCACCTTTAAAAAATATGATACATCAAGTGCTTTAGGTAG
 E Q E R Y R I I S L S V I S S L I S K I 2880
 AGAACAGGAAAGGTACAGGATTATATCCTTGCTGTATTTCAGTTTGATTAGTAAAT
 L S L L S L I L T V S L T L P Y L G Q E 2940
 ACTCTCACTACTTTCTCTTATATTAAGTGTAAAGTTTAACTTTACCTTATTTAGGACAAGA
 R F G V W M T I T S L G A A L T F L D L 3000
 GAGATTTGGTGTATGGATGACTATTACCAGTCTTGGTGCTGCTCTGACATTTTGGACTT
 G I G N A L T N R I A H S F A C G K N L 3060
 AGGTATAGGAAATGCATTAACAAACAGGATCGCACATTCATTTGCGTGTGGCAAAATTT
 K M S R Q I S G G L T L L A G L S F V I 3120
 AAAGATGAGTCGGCAAATTAGTGGTGGGCTCACTTTGCTGGCTGGATTATCGTTTGTCTAT
 T A I C Y I T S G M I D W Q L V I K G I 3180
 AACTGCAATATGCTATATTACTTCTGGCATGATTGATTGGCAACTAGTAATAAAAGGTAT
 N E N V Y A E L Q H S I K V F V I I F G 3240
 AAACGAGAATGTGTATGCAGAGTTACAACACTCAATTAAAGTCTTTGTAATCATATTTGG
 L G I Y S N G V Q K V Y M G I Q K A Y I 3300
 ACTTGGAATTTATTCAAATGGTGTGCAAAAAGTTTATATGGGAATACAAAAGCCTATAT
 S N I V N A I F I L L S I I T L V I S S 3360
 AAGTAATATTGTTAATGCCATATTTATATTGTTATCTATTATTACTCTAGTAATATCGTC
 K L H A G L P V L I V S T L G I Q Y I S 3420
 GAAACTACATGCGGGACTACCAGTTTAAATTGTCAGCACTCTTGGTATTCAATACATATC
 G I Y L T I N L I I K R L I K F T K V N 3480
 GGGAACTCTATTTAAACAATTAATCTTATTATAAAGCGATTAATAAAGTTTACAAAAGTTAA
 I H A K R E A P Y L I L N G F F F F I L 3540
 CATACATGCTAAAAGAGAAGCTCCATATTTGATATTAAACGGTTTTTTCTTTTTTATTTT
 Q L G T L A T W S G D N F I I S I T L G 3600
 ACAGTTAGGCACTCTGGCAACATGGAGTGGTGATAACTTTATAATATCTATAACATTGGG

20/96

V T Y V A V F S I T Q R L F Q I S T V P
TGTTACTTATGTTGCTGTTTTTAGCATTACACAGAGATTATTTCAAATATCTACGGTCCC 3660

L T I Y N I P L W A A Y A D A H A R N D
TCTTACGATTTATAACATCCCCTTATGGGCTGCTTATGCAGATGCTCATGCACGCAATGA 3720

T Q F I K K T L R T S L K I V G I S S F
TACTCAATTTATAAAAAAGACGCTCAGAACATCATTGAAAATAGTGGGTATTTTCATCATT 3780

L L A F I L V V F G S E V V N I W T E G
CTTATTGGCCTTCATATTAGTAGTGTTCGGTAGTGAAGTCGTTAATATTTGGACAGAAGG 3840

K I Q V P R T F I I A Y A L W S V I D A
AAAGATTTCAGGTACCTCGAACATTCATAATAGCTTATGCTTTTATGGTCTGTTATTGATGC 3900

F S N T F A S F L N G L N I V K Q Q M L
TTTTTCGAATACATTTGCAAGCTTTTTTAAATGGTTTGAACATAGTTAAACAACAAATGCT 3960

A V V T L I L I A I P A K Y I I V S H F
TGCTGTTGTAACATTGATATTGATCGCAATTCCAGCAAAATACATCATAGTTAGCCATTT 4020

G L T V M L Y C F I F I Y I V N Y F I W
TGGGTTAACTGTTATGTTGTAAGTCTGCTTCATTTTATATATATTTGTAATTTACTTTATATG 4080

Start of orf5, End of orf4
M K M

Y K C S F K K H I D R Q L N I R G *
GTATAAATGTAGTTTTTAAAAAACATATCGATAGACAGTTAAATATAAGAGGATGAAAATG 4140

K Y I P V Y Q P S L T G K E K E Y V N E
AAATATATACCAAGTTTACCAACCGTCATTGACAGGAAAGAAAGAAATATGTAATGAA 4200

C L D S T W I S S K G N Y I Q K F E N K
TGTCTGGAAGTCAAGCTGGATTTCATCAAAAGGAACTATATTTCAGAACTTTGAAAATAAA 4260

F A E Q N H V Q Y A T T V S N G T V A L
TTTGCGGAACAAAACCATGTGCAATATGCAACTACTGTAAAGTAATGGAAACGGTTGCTCTT 4320

H L A L L A L G I S E G D E V I V P T L
CATTTAGCTPTTGTAGCGTTAGGTATATCGGAAGGAGATGAAGTTATTGTTCCAAACCTG 4380

T Y I A S V N A I K Y T G A T P I F V D
ACATATATAGCATCACTTAATGCTATATAAATACACAGGAGCCACCCCCATTTTGGTTGAT 4440

S D N E T W Q M S V S D I E Q K I T N K
TCAGATAATGAACTTGGCAATGTCTGTAGTGAACATAGAACAAAAATCACTAATAAA 4500

A T K A I M C V H L Y G H P C D M E Q I V
ACTAAAGCTATTATGTGTGTCCATTTATACGGACATCCATGTGATATGGAACAAATTGTA 4560

E L A K S R N L F V I E D C A E A F G S
GAAGTGGCCAAAAGTAGAAATTTGTTTGTAAATTGAAGATTGCGCTGAAGCCTTTGCTTCT 4620

K Y K G K Y V G T F G D I S T F S F F G
AAATATAAGGTAAATATGTGGGAACATTTGGAGATATTTCTACTTTTATAGCTTTTTTGG 4680

N K T I T T G E G G M V V T N D K T L Y
AATAAACTATTACTACAGGTGAAGGTGGAAATGGTTGTCAAGAAATGACAAAACACTTTAT 4740

D R C L H F K G Q G L A V H R Q Y W H D
GACCGTTGTTTTACATTTTAAAGGCCAAGGATTAGCTGTACATAGGCAATATTGGCATGAC 4800

V I G Y N Y R M T N I C A A I G L A Q L
GTTATAGGCTACAATTATAGGATGACAAATATCTGCGCTGCTATAGGATTAGCCCCAGTTA 4860

21/96

E Q A D D F I S R K R E I A D I Y K K N 4920
 GAACAAGCTGATGATTTTATATCAGGAAAACCTGAAATTGCTGATATTTATAAAAAAAT
 I N S L V Q V H K E S K D V F H T Y W M 4980
 ATCAACAGTCTTCTACAAAGTCCACAAAGGAAAGTAAAGATGTTTTTCACACTTATTGGATG
 V S I L T R T A E E R E E L R N H L A D 5040
 GTCTCAATTCTAACTAGGACCGCAGAGGAAAGAGAGGAATTAAGGAATCACCTTGCAGAT
 K L I E T R P V F Y P V H T M P M Y S E 5100
 AAATCATCGAAACAAAGGCCAGTTTTTTACCCTGTCCACACGATGCCAATGTACTCCGAA
 K Y Q K H P I A E D L G W R G I N L P S 5160
 AAATATCAAAAGCACCTATAGCTGAGGATCTTGGTTGGCTGGAAATTAATTTACCTAGT
 F P S L S N E Q V I Y I C E S I N E F Y 5220
 TTECCCCAGCCTATCGAATGAGCAAGTTATTTATATTTGTGAATCTATTAACGAATTTTAT

End of orf5 **Start of orf6**
 S D K * M K I A L N S D 5280
 AGTGATAAATAGCCTAAATATTTGTAAAGTCAATTCATGAAATTCGCTTGAATTCAGAT
 G F Y E W G G G I D F I K Y I L S I L E 5340
 GGATTTTACGAGTGGGGCGGTGGAATTTGATTTTATTAAATATATTCTGTCAATATTAGAA
 T K P E I C I D I L L P R N D I H S L I 5400
 ACGAAACCAGAAATATGTATCGATATTCTTTACCGAGAAATGATATACATTCTCTTATA
 R E K A F P F K S I L K A I L K R E R P 5460
 AGAGAAAAAGCATTTCCTTTTAAAAGTATATTAAAAGCAATTTTAAAGAGGGAAAGCCCT
 R W I S L N R F N E Q Y Y R D A F T Q N 5520
 CGATGGATTTTCATTAAATAGATTTAATGAGCAATACTATAGAGATGCCTTTACACAAAAT
 N I E T N L T F I K S K S S A F Y S Y F 5580
 AATATAGAGACGAATCTTACCTTTATTAAAAGTAAGAGCTCTGCCTTTTATTTCATATTTT
 D S S D C D V I L P C M R V P S G N L N 5640
 GATAGTAGCGATTGTGATGTTATTCTTCCTTGCATGCGTGTTCCTTCGGGAAATTTGAAT
 K K A W I G Y I Y D F Q H C Y Y P S F F 5700
 AAAAAAGCATGGATTGGTTATATTATGACTTTCAACACTGTTACTATCCTTCATTTTTT
 S K R E I D Q R N V F F K L M L N C A N 5760
 AGTAAGCGAGAAATAGATCAAAGGAATGTGTTTTTTAAATTGATGCTCAATTGCGCTAAC
 N I I V N A H S V I T D A N K Y V G N Y 5820
 AATATTATTGTTAATGCACATTCAGTTATTACCGATGCAATAAATATGTTGGGAATTAT
 S A K L H S L P F S P C P Q L K W F A D 5880
 TCTGCAAAACTACATTCTCTCCATTTAGTCCATGCCCTCAATTAAAATGGTTCGCTGAT
 Y S G N I A K Y N I D K D Y F I I C N Q 5940
 TACTCTGGTAATATTGCCAAATATAATATTGACAAGGATTATTTTATAATTTGCAATCAA
 F W K H K D H A T A F R A F K I Y T E Y 6000
 TTTTGGAACATAAAGATCATGCAACTGCTTTTAGGGCATTTAAAATTTTACTGAATAT
 N P D V Y L V C T G A T Q D Y R F P G Y 6060
 AATCCTGATGTTTATTAGTATGCACGGGAGCTACTCAAGATTATCGATTCCCTGGATAT
 F N E L M V L A K K L G I E S K I K I L 6120
 TTTAATGAATTGATGGTTTTGGCAAAAAGCTCGGAATTGAATCGAAAATTAAGATATTA

22/96

G H I P K L E Q I E L I K N C I A V I Q
 GGGCATATACCTAAACTTGAACAAATTGAATTAATCAAAAATTGCATTGCTGTAATACAA 6180
 P T L F E G G P G G G V T F D A I A L G
 CCAACCTTATTTGAAGGCGGGCTGGAGGGGGGTAACATTTGACGCTATTGCATTAGGG 6240
 K K V I L S D I D V N K E V N C G D V Y
 AAAAAAGTTATACTATCTGACATAGATGTCAATAAAGAAGTTAATTGCGGTGATGTATAT 6300
 F F Q A K N H Y S L N D A M V K A D E S
 TTCTTTTCAGGCAAAAACCATTTATTCATTAAATGACGCGATGGTAAAAGCTGATGAATCT 6360
 K I F Y E P T T L I E L G L K R R N A C
 AAAATTTTTTATGAACCTACAACCTCTGATAGAATTGGGTCTCAAAAGACGCAATGCGTGT 6420
 A D F L L D V V K Q E I E S R S *
 GCAGATTTTCTTTTAGATGTTGTGAAACAAGAAATTGAATCCCGATCT TAATATATTCAA 6480
 End of orf6
 Start of orf7
 M T K V A L I T G V T G Q D G S Y
 GAGGTATATAATGACTAAAGTCGCTCTTATTACAGGTGTAAGTGGACAAGATGGATCTTA 6540
 L A E F L L D K G Y E V H G I K R R A S
 TCTAGCTGAGTTTTTGCTTGATAAAGGGTATGAAGTTCATGGTATCAAACGCCGAGCCTC 6600
 S F N T E R I D H I Y Q D P H G S N P N
 ATCTTTTAATACAGAACGCATAGACCATATTTATCAAGATCCACATGGTTCTAACCCAAA 6660
 F H L H Y G D L T D S S N L T R I L K E
 TTTTCACTTGCACTATGGAGATCTGACTGATTCATCTAACCTCACTAGAATTCTAAAGGA 6720
 V Q P D E V Y N L A A M S H V A V S F E
 GGTACAGCCAGATGAAGTATATAATTTAGCTGCTATGAGTCACGTAGCAGTTTCTTTTGA 6780
 S P E Y T A D V D A I G T L R L L E A I
 GTCTCCAGAATATACAGCCGATGTGATGCAATTGGTACATTACGTTTACTGGAAGCAAT 6840
 R F L G L E N K T R F Y Q A S T S E L Y
 TCGCTTTTTTAGGATTGGAAAACAAAACGCGTTTCTATCAAGCTTCAACCTCAGAATTATA 6900
 G L V Q E I P Q K E S T P F Y P R S P Y
 TGGACTTGTTTCAGGAAATCCCTCAAAAAGAATCCACCCCTTTTTATCCTCGTTCCCTTA 6960
 A V A K L Y A Y W I T V N Y R E S Y G I
 TGCAGTTGCAAACTTTACGCATATTGGATCACGGTAAATTATCGAGAGTCATATGGTAT 7020
 Y A C N G I L F N H E S P R R G E T F V
 TTATGCATGTAATGGTATATTGTTCAATCATGAATCTCCACGCCGTGGAGAAACGTTTGT 7080
 T R K I T R G L A N I A Q G L E S C L Y
 AACAAGGAAAATTACTCGAGGACTTGCAATATTGCACAAGGCTTGGAATCATGTTTGT 7140
 L G N M D S L R D W G H A K D Y V R M Q
 TTTAGGGAATATGGATTTCGTTACGAGATTGGGGACATGCAAAAGATTATGTTAGAATGCA 7200
 W L M L Q Q E Q P E D F V I A T G V Q Y
 ATGGTTGATGTTACAACAGGAGCAACCCGAAGATTTTGTGATTGCAACAGGAGTCCAATA 7260
 S V R Q F V E M A A A Q L G I K M S F V
 CTCAGTCCGTCAGTTTGTGCAATGGCAGCAGCACAACTTGGTATTAAGATGAGCTTTGT 7320

23/96

G K G I E E K G I V D S V E G Q D A P G
 TGGTAAAGGAATCGAAGAAAAAGGCATTGTAGATTTCGGTTGAAGGACAGGATGCTCCAGG 7380

V K P G D V I V A V D P R Y F R P A E V
 TGTGAAACCAGGTGATGTCATTGTTGCTGTTGATCCTCGTTATTTCCGACCAGCTGAAGT 7440

D T L L G D P S K A N L K L G W R P E I
 TGATACTTTGCTTGAGATCCGAGCAAAGCTAATCTCAAACCTTGGTTGGAGACCAGAAAT 7500

T L A E M I S E M V A K D L E A A K K H
 TACTCTTGCTGAAATGATTTCTGAAATGGTTGCCAAAGATCTTGAAGCCGCTAAAAACA 7560

Start of orf8, End of orf7

S L L K S H G F S V S L A L E * M M M N K
 TTCTCTTTTAAATCGCATGGTTTTTCTGTAAGCTTAGCTCTGGAATGATGATGAATAAG 7620

Q R I F I A G H Q G M V G S A I T R R L
 CAACGTATTTTATTTGCTGGTCACCAAGGAATGGTTGGATCAGCTATTACCCGACGCCTC 7680

K Q R D D V E L V L R T R D E L N L L D
 AAACAACGTGATGATGTTGAGTTGGTTTTACGTACTCGGGATGAATTGAACTTGTGGAT 7740

S S A V L D F F S S Q K I D Q V Y L A A
 AGTAGCGCTGTTTTGGATTTTTTTCTTCACAGAAAATCGACCAGGTTTATTTGGCAGCA 7800

A K V G G I L A N S S Y P A D F I Y E N
 GCAAAAGTCGGAGGTATTTTAGCTAACAGTTCTTATCCTGCCGATTTTATATATGAGAAT 7860

I M I E A N V I H A A H K N N V N K L L
 ATAATGATAGAGGCGAATGTCATTCATGCTGCCACAAAAATAATGTAAATAAACTGCTT 7920

F L G S S C I Y P K L A H Q P I M E D E
 TTCCTCGGTTTCGTCGTGTATTTATCCTAAGTTAGCACACCAACCGATTATGGAAGACGAA 7980

L L Q G K L E P T N E P Y A I A K I A G
 TTATTACAAGGGAAACTTGAGCCAACAAATGAACCTTATGCTATCGCAAAAATTGCAGGT 8040

I K L C E S Y N R Q F G R D Y R S V M P
 ATTAAATTATGTGAATCTTATAACCGTCAGTTTGGGCGTGATTACCGTTCAGTAATGCCA 8100

T N L Y G P N D N F H P S N S H V I P A
 ACCAATCTTTATGGTCCAAATGACAATTTTCATCCAAGTAATTCTCATGTGATTCGGCG 8160

L L R R F H D A V E N N S P N V V V W G
 CTTTTGCGCCGCTTTCATGATGCTGTGGAACAATTCTCCGAATGTTGTTGTTGGGGA 8220

S G T P K R E F L H V D D M A S A S I Y
 AGTGGTACTCCAAAGCGTGAATTCTTACATGTAGATGATATGGCTTCTGCAAGCATTTAT 8280

V M E M P Y D I W Q K N T K V M L S H I
 GTCATGGAGATGCCATACGATATATGGCAAAAAATACTAAAGTAATGTTGTCTCATATC 8340

N I G T G I D C T I C E L A E T I A K V
 AATATTGGAACAGGTATTGACTGCACGATTTGTGAGCTTGCGGAAACAATAGCAAAAGTT 8400

V G Y K G H I T F D T T K P D G A P R K
 GTAGGTTATAAAGGGCATATTACGTTGATACAACAAAGCCCGATGGAGCCCCCTCGAAAA 8460

L L D V T L L H Q L G W N H K I T L H K
 CTACTTGATGTAACGCTTCTTCATCAACTAGGTTGGAATCATAAAATTACCCCTCACAAG 8520

24/96

End of orf8

G L E N T Y N W F L E N Q L Q Y R G *
 GGTCTTGAAAATACATACAACCTGGTTTCTTGAAAACCAACTTCAATATCGGGGG TAATAA 8580

Start of orf9
 M F L H S Q D F A T I V R S T P L I S I
 TGTTTTTACATTCCCAAGACTTTGCCACAATTGTAAGGTCTACTCCTCTTATTTCTATAG 8640

D L I V E N E F G E I L L G K R I N R P
 ATTTGATTGTGGAACGAGTTTGGCGAAATTTTGCTAGGAAAACGAATCAACCGCCCCG 8700

A Q G Y W F V P G G R V L K D E K L Q T
 CACAGGGCTATTGGTTCGTTCCTGGTGGTAGGGTGTGAAAGATGAAAATTGCAGACAG 8760

A F E R L T E I E L G I R L P L S V G K
 CCTTTGAACGATTGACAGAAATTGAACAGGAATTCGTTTGCCTCTCTCTGTGGGTAAGT 8820

F Y G I W Q H F Y E D N S M G G D F S T
 TTTATGGTATCTGGCAGCACTTCTACGAAGACAATAGTATGGGGGGAGACTTTTCAACGC 8880

H Y I V I A F L L K L Q P N I L K L P K
 ATTATATAGTTATAGCATTCTTCTTAAATTACAACCAAACATTTTGAAATTACCGAAGT 8940

S Q H N A Y C W L S R A K L I N D D D V
 CACAACATAATGCTTATTGCTGGCTATCGCGAGCAAAGCTGATAAATGATGACGATGTGC 9000

H Y N C R A Y F N N K T N D A I G L D N
 ATTATAATTGTCGCGCATATTTTAACAATAAAACAAATGATGCGATTGGCTTAGATAATA 9060

Start of orf10 End of orf9

M S D A P I I A V V M A G G T G S
 K D I I C L M R Q *
 AGGATATAATATGCTCTGATGCGCCAATAATTGCTGTAGTTATGGCCGGTGGTACAGGCAG 9120

R L W P L S R E L Y P K Q F L Q L S G D
 TCGTCTTTGGCCACTTTCTCGTGAACCTATATCCAAAGCAGTTTTTACAACCTCTCTGGTGA 9180

N T L L Q T T L L R L S G L S C Q K P L
 TAACACCTTGTTACAAACGACTTTGCTACGACTTTTACGGCCTATCATGTCAAAAACCATT 9240

V I T N E Q H R F V V A E Q L R E I N K
 AGTGATAACAAATGAACAGCATCGCTTTGTTGTGGCTGAACAGTTAAGGGAAATAAATAA 9300

L N G N I I L E P C G R N T A P A I A I
 ATTAAATGGTAATATTATTCTAGAACCATGCGGGCGAAATACTGCACCAGCAATAGCGAT 9360

S A F H A L K R N P Q E D P L L L V L A
 ATCTGCGTTTCATGCGTTAAAACGTAATCCTCAGGAAGATCCATTGCTTCTAGTTCTTGC 9420

A D H V I A K E S V F C D A I K N A T P
 GGCAGACCACGTTATAGCTAAAGAAAGTGTTTCTGTGATGCTATTAAAAATGCAACTCC 9480

I A N Q G K I V T F G I I P E Y A E T G
 CATCGCTAATCAAGGTAAAATTGTAACGTTTGGAAATTATACCAGAATATGCTGAAACTGG 9540

Y G Y I E R G E L S V P L Q G H E N T G
 TTATGGGTATATTGAGAGAGGTGAACCTATCTGTACCGCTTCAAGGGCATGAAAATACTGG 9600

F Y Y V N K F V E K P N R E T A E L Y M
 TTTTTATTATGTAAATAAGTTTGTGCGAAAAGCCTAATCGTGAAACCGCAGAATTGTATAT 9660

T S G N H Y W N S G I F M F K A S V Y L
 GACTTCTGGTAATCACTATTGGAATAGTGAATATTTCATGTTTAAGGCATCTGTTTATCT 9720

25/96

E E L R K F R P D I Y N V C E Q V A S S
 TGAGGAATTGAGAAAATTTAGACCTGACATTTACAATGTTTGTGAACAGGTTGCCCTCATC 9780
 S Y I D L D F I R L S K E Q F Q D C P A
 CTCATACATTGATCTAGATTTTATTTCGATTATCAAAAAGAACAATTTCAAGATTGTCCTGC 9840
 E S I D F A V M E K T E K C V V C P V D
 TGAATCTATTGATTTTCTGCTGTAATGGAAAAACAGAAAAATGTGTTGTATGCCCTGTTGA 9900
 I G W S D V G S W Q S L W D I S L K S K
 TATTGGTTGGAGTGACGTTGGATCTTGGCAATCGTTATGGGACATTAGTCTAAAATCGAA 9960
 T G D V C K G D I L T Y D T K N N Y I Y
 AACAGGAGATGTATGTAAAGGTGATATATTAACCTATGATACTAAGAATAATTATATCTA 10020
 S E S A L V A A I G I E D M V I V Q T K
 CTCTGAGTCAGCGTTGGTAGCCGCCATTGGAATTGAAGATATGGTTATCGTGCAAACATAA 10080
 D A V L V S K K S D V Q H V K K I V E M
 AGATGCCGTTCTTGTGTCTAAAAAGAGTGATGTACAGCATGTAAAAAAAATAGTCGAAAT 10140
 L K L Q Q R T E Y I S H R E V F R P W G
 GCTTAAATTGCAGCAACGTACAGAGTATATTAGTCATCGTGAAGTTTCCGACCATGGGG 10200
 K F D S I D Q G E R Y K V K K I I V K P
 AAAATTTGATTGCGATTGACCAAGGTGAGCGATACAAAGTCAAGAAAATTATTGTGAAACC 10260
 G E G L S L R M H H H R S E H W I V L S
 TGGTGAGGGGCTTTCTTTAAGGATGCATCACCATCGTTCTGAACATTGGATCGTGCTTTC 10320
 G T A K V T L G D K T K L V T A N E S I
 TGGTACAGCAAAAGTAACCCCTTGGCGATAAACTAACTAGTCACCGCAAATGAATCGAT 10380
 Y I P L G A A Y S L E N P G I I P L N L
 ATACATTTCCCTTGGCGCAGCGTATAGTCTTGAGAATCCGGGCATAATCCCTCTTAATCT 10440
 I E V S S G D Y L G E D D I I R Q K E R
 TATTGAAGTCAGTTTCAAGGGATTATTTGGGAGAGGATGATATTATAAGACAGAAAGAACG 10500
 End of orf10 Start of orf11
 Y K H E D * M K S L T C F K A Y D I R
 TTACAAACATGAAGATTAACATATGAAATCTTTAACCTGCTTTAAAGCCTATGATATTTCG 10560
 G K L G E E L N E D I A W R I G R A Y G
 CGGGAAATTAGGCGAAGAACTGAATGAAGATATTGCCTGGCGCATTTGGGCGTGCCTATGG 10620
 E F L K P K T I V L G G D V R L T S E A
 CGAATTTCTCAAACCGAAAACCATTTGTTTTAGGCGGTGATGTCCGCCTCACCAGCGAAGC 10680
 L K L A L A K G L Q D A G V D V L D I G
 GTTAAACTGGCGCTTGCGAAAGGTTTACAGGATGCGGGCGTCGATGTGCTGGATATCGG 10740
 M S G T E E I Y F A T F H L G V D G G I
 TATGTCCGGCACCGAAGAGATCTATTTGCCACGTTCCATCTCGGAGTGGATGGCGGCAT 10800
 E V T A S H N P M D Y N G M K L V R E G
 CGAAGTTACCGCCAGCCATAACCCGATGGATTACAACGGCATGAAGCTGGTGCGCAAGG 10860
 A R P I S G D T G L R D V Q R L A E A N
 GGCTCGCCCGATCAGCGGTGATACCGGACTGCGCGATGTCCAGCGTCTGGCAGAAGCCAA 10920
 D F P P V D E T K R G R Y Q Q I N L R D
 TGACTTCCCTCCTGTCGATGAAACCAACGTGGTCGCTATCAGCAAATCAATCTGCGTGA 10980

26/96

A Y V D H L F G Y I N V K N L T P L K L
CGCTTACGTTGATCACCTGTTTCGGTTATATCAACGTCAAAAACCTCACGCCGCTCAAGCT 11040

V I N S G N G A A G P V V D A I E A R F
GGTGATCAACTCCGGGAACGGCGCAGCGGGTCCGGTGGTGGACGCCATTGAAGCCCCGATT 11100

K A L G A P V E L I K V H N T P D G N F
TAAAGCCCTCGGCGCACCGGTGGAATTAATCAAAGTACACAACACGCCGGACGGCAATTT 11160

P N G I P N P L L P E C R D D T R N A V
CCCCAACGGTATTCCCTAACCCGCTGCTGCCGGAATGCCGCGACGACACCCGTAATGCGGT 11220

I K H G A D M G I A F D G D F D R C F L
CATCAACACGGCGCGGATATGGGCATTGCCTTTGATGGCGATTTTGACCGCTGTTTCCT 11280

F D E K G Q F I E G Y Y I V G L L A E A
GTTTGACGAAAAAGGGCAGTTTATCGAGGGCTACTACATTGTGCGCCTGCTGGCAGAAGC 11340

F L E K N P G A K I I H D P R L S W N T
GTTCCCTCGAAAAAATCCCGGCGCGAAGATCATCCACGATCCACGTCTCTCCTGGAACAC 11400

V D V V T A A G G T P V M S K T G H A F
CGTTGATGTGGTGACTGCCGCGAGGCGGCACCCCGGTAATGTGCGAAAACCGGACACGCCCTT 11460

I K E R M R K E D A I Y G G E M S A H H
TATTAAAGAACGTATGCGCAAGGAAGACGCCATCTACGGTGGCGAAATGAGCGCTCACCA 11520

Y F R D F A Y C D S G M I P W L L V A E
TTACTTCCGTGATTTTCGCTTACTGCGACAGCGGCATGATCCCGTGGCTGCTGGTCGCCGA 11580

L V C L K G K T L G E M V R D R M A A F
ACTGGTGTGCCTGAAAGGAAAAACGCTGGGCGAAATGGTGCGCGACCGGATGGCGGCGTT 11640

P A S G E I N S K L A Q P V E A I N R V
TCCGGCAAGCGGTGAGATCAACAGCAAACCTGGCGCAACCCGTTGAGGCAATTAATCGCGT 11700

E Q H F S R E A L A V D R T D G I S M T
GGAACAGCATTTTAGCCGCGAGGCGCTGGCGGTGGATCGCACCGATGGCATCAGCATGAC 11760

F A D W R F N L R S S N T E P V V R L N
CTTTGCCGACTGGCGCTTTAACCTGCGCTCCTCCAACACCGAACCGGTGGTGCGGTTGAA 11820

V E S R G D V K L M E K K T K A L L K L
TGTGGAATCACGCGGTGATGTAAAGCTAATGGAAAAGAAAATAAAGCTCTTCTTAAATT 11880

End of orf11

L S E *
GCTAAGTGAGTGATTATTTACATTAATCATTAAGCGTATTTAAGATTATATTAAAGTAAT 11940

GTTATTGCGGTATATGATGAATATGTGGGCTTTTTTATGTATAACGACTATACCGCAACT 12000

Start of H-repeat

TTATCTAGGAAAAAGATTAATAGAAATAAAGTTTTGTACTGACCAATTTGCATTTTCACGTC 12060

ACGATTGAGACGTTCCCTTTGCTTAAGACATTTTTTTCATCGCTTATGTAATAACAAATGTG 12120

CCTTATATAAAAAAGGAGAACAAAATGGAACCTAAATAATTGAGACAATAGATTTTTTATT 12180

ATCCCTGTTTACGATATTATAGCCAAAGTTGTATCCTGCATCAGTCCTGCAATATTTTAC 12240

GAGTGCTTTGTTAACTGAATACATGTCTGCCATTTTCCAGATGATAACGACGTCATCGCA 12300

ATTGATGGTAAACACTTCGGCACACTTATGACAAGAGTCGTCGAGAGGAGTGGTTCAT 12360

27/96

GTCATTAGTGCCTTTCAGCAATGCACAGTCTGGTCCCTCGGATAGATCAAGACGGATGAGA 12420
 AACCTAATGCGTTCACAGTTATTCATGAACTTTCTAAAATGATGGGTATTAAAGGAAAAA 12480
 TAATCATAACTGATGCGATGGCTTGCCAGAAAAGATATTGCAGAGAAGATATAAAAACAGA 12540
 GATGTGATTATTTATTCGCTGTAAAAGGAAATAAGAGTCCGGCTTAATAGAGTCTTTGAGG 12600
 AGATATTTACGCTGAAAGAATTAAATAATCCAAAACATGACAGTTACGCAATTAGTGAAA 12660
 AGAGGCACGGCAGAGACGATGTCCGTCTTCATATTGTTTGAGATGCTCCTGATGAGCTTA 12720
 TTGATTTACGTTTGAATGGAAAGGGCTGCAGAAATTTATGAATGGCAGTCCACTTTCTCT 12780
 CAATAATAGCAGAGCAAAAGAAAGAATCCGAAATGACGATCAAAATATTATATTAGATCTG 12840
 CTGCTTTAACCGCAGAGAAGTTCGCCACAGTAAATCGAAATCACTGGCGCATGGAGAATA 12900
 AGTTGCACAGTAGCCTGATGTGGTAATGAATGAAATCGACTATAATATAAGAAGGCGAGT 12960
 TGCATTGCAATGATTTTCTAGAATGCGGCACATCGCTATTAATATCTGACAATGATAATG 13020
 TATTCAAGGCAGGATTATCATGTAAGATGCGAAAAGCAGTCATGGACAGAAACTTCCTAG 13080
 CGTCAGGCATTGCAGCGTGCAGGGCTTTCATAATCTTGCAAT **End of the H-repeat** 13140
 TGGTTTGTGATAAGATATTTTC
Start of orf12
 M N L Y G I F G A G S Y G R E
 TTTGGAGATGGGAAAATGAATTTGTATGGTATTTTGGTGCTGGAAGTTATGGTAGAGAA 13200
 T I P I L N Q Q I K Q E C G S D Y A L V
 ACAATACCCATTCTAAATCAACAAATAAAGCAAGAATGTGGTTCTGACTATGCTCTGGTT 13260
 F V D D V L A G K K V N G F E V L S T N
 TTTGTGGATGATGTTTTGGCAGGAAAGAAAGTTAATGGTTTTGAAGTGCTTTCAACCAAC 13320
 C F L K A P Y L K K Y F N V A I A N D K
 TGCTTTCTAAAAGCCCCTTATTTAAAAAAGTATTTTAATGTTGCTATTGCTAATGATAAG 13380
 I R Q R V S E S I L L H G V E P I T I K
 ATACGACAGAGAGTGTCTGAGTCAATATTATTACACGGGGTTGAACCAATAACTATAAAA 13440
 H P N S V V Y D H T M I G S G A I I S P
 CATCCAAATAGCGTTGTTTATGATCATACTATGATAGGTAGTGGCGCTATTATTTCTCCC 13500
 F V T I S T N T H I G R F F H A N I Y S
 TTTGTTACAATATCTACTAATACTCATATAGGGAGGTTTTTTCATGCAAACATATACTCA 13560
 Y V A H D C Q I G D Y V T F A P G A K C
 TACGTTGCACATGATTGTCAAATAGGAGACTATGTTACATTTGCTCCTGGGGCTAAATGT 13620
 N G Y V V I E D N A Y I G S G A V I K Q
 AATGGATATGTTGTTATTGAAGACAATGCATATATAGGCTCGGGTGCAGTAATTAAGCAG 13680
 G V P N R P L I I G A G A I I G M G A V
 GGTGTTCCATAATCGCCCACTTATTATTGGCGCGGGAGCCATTATAGGTATGGGGGCTGTT 13740
 V T K S V P A G I T V C G N P A R E M K
 GTCATAAAAGTGTTCCTGCCGTATAACTGTGTGCGGAAATCCAGCAAGAGAAATGAAA 13800
End of orf12
 R S P T S I *
 AGATCGCCAACATCTATT TAATGGGAATGCGAAAACACGTTCCAAATGGGACTAATGTTT 13860

28/96

AAAATATATATAATTTTCGCTAATTTACTAAATTATGGCTTCTTTTAAAGCTATCCTTTAC 13920
TTAGTTATTACTGATACAGCATGAAATTTATAATACTCTGATACATTTTATACGTTATT 13980
CAAGCCGCATATCTAGCGGTAACCCCTGACAGGAGTAAACAATG 14024

29/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAATAATATCAACAAG
AACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAACAGC
GCGAAGGATGACGCCCGGGTCAGGCGATTGCTAACCGTTTTACTTCTAACATTAAAGGC
CTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCACAGACCACTGAAGGC
GCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGAGCTGACGGTTCAGGCTTCT
ACCGGGACTAACTCTGATTCCGATCTGGACTCCATTACAGGACGAAATCAAATCCCGTCTC
GACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACGTACTGGCAAAA
GACGGTTCGATGAAAAATTCAGGTAGGTGCGAACGACGGCCAGACTATCACTATTGATCTG
AAGAAAATTGACTCTGATACGCTGGGGCTGAATGGTTTTAACGTGAATGGTTCGGGTACG
ATAGCCAATAAAGCGGCGACCATTAGCGACCTGACAGCAGCGAAAATGGATGCTGCAACT
AATACTATAACTACAACAAATAATGCGCTGACTGCATCAAAGGCCCTTGATCACTGAAA
GATGGTGACACTGTTACTATCAAAGCAGATGCAGCTCAAACCTGCCACGGTCTATACATAC
AATGCATCTGCTGGTAACTTCTCATTAGTAATGTATCGAATAATACTTCAGCAAAAGCA
GGTGATGTAGCAGCTAGCCTTCTCCCGCCGGCTGGGCAAACTGCTAGTGGTGTTTACAAA
GCAGCAAGCGGTGAAGTGAACCTTTGATGTTGATGCGAATGGTAAAATTACAATCGGAGGA
CAGGAAGCCTATTTAACTAGTGATGGTAACTTAACTACAAACGATGCTGGTGGTGCGACT
GCGGCTACGCTTGATGGTTTTATTCAAGAAAGCTGGTGATGGTCAATCAATCGGGTTTAAT
AAGACTGCATCAGTCACGATGGGGGGAACAACCTTATAACCTTTAAACGGGTGCTGATGCT
GGTGCTGCAACTGCTAACGCAGGGGTATCGTTCACCTGATACAGCTAGCAAAGAAACCGTT
TTAAATAAAGTGGCTACAGCTAAACAAGGCACAGCAGTTGCAGCTAACGGTGATACATCC
GCAACAATTACCTATAAATCTGGCGTTTACAGCGTATCAGGCGGTATTTGCCGAGGTGAC
GGTACTGCTAGCGCAAAATATGCCGATAATACTGACGTTTCTAATGCAACAGCAACATAC
ACAGATGCTGATGGTGAAATGACTACAATTGGTTCATACACCACGAAGTATTCAATCGAT
GCTAACAAACGGCAAGGTAACGTGTTGATTCTGGAACCTGGTTCGGGTAAATATGCGCCGAAA
GTCGGGGCTGAAGTATATGTTAGTGCTAATGGTACTTTAACAACAGATGCAACTAGCGAA
GGCACAGTAACAAAAGATCCACTGAAAGCTCTGGATGAAGCTATCAGCTCCATCGACAAA
TTCCGTTTCATCCCTGGGGGCTATCCAAAACCGTTTGGATTCCGCCGTCACCAACCTGAAC
AACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTACAGGACGCCGACTATGCGACC
GAAGTGTCACCATGTGCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCA
AAAGCCAACCAGGTACCGCAGCAGGTTCTGTCTCTACTGCAGGGTTAA

Figure 7

30/96

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATT
GAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGGCG
ATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAAAT
GATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTG
CAGCGTATTCGTGAACTTTCTGTTCAGGCAACTAACGGTACTAACTCTGACAGTGACCTG
ACCTCCATCCAGTCCGAAATCCAGCAGCGTCTGAGTGAAATTGACCGTGTCTTCTGGTCAG
ACTCAGTTTAAACGGCGTTAAAGTGCTGGCTTCTGATCAGGATATGACTATTCAGGTTGGT
GCAAACGACGGCGAAACAATTACTATTAAACTGCAGGAAATTAATTCCGACACACTGGGA
TTATCTGGTTTTTGGTATTAAAGATCCTACTAAATTAAAAGCCGCAACGGCTGAAACAACC
TATTTTGGATCGACAGTTAAGCTTGCTGACGCTAATACACTTGATGCAGATATTACAGCT
ACAGTTAAAGGCACTACGACTCCGGGCCAACGTGACGGTAATATTATGTCTGATGCTAAC
GGTAAGTTGTACGTTAAAGTTGCCGGTTCAGATAAACCCGCTGAAAATGGTTATTATGAA
GTTACTGTGGAGGATGATCCGACATCTCCTGATGCAGGTAAGCTGAAGCTGGGGGCTCTA
GCGGGTACCCAGCCTCAAGCTGGTAATTTAAAGGAAGTCACAACGGTGAAAGGGAAGGGG
GCTATTGATGTTTCAAGTTGGGTACTGATACCGCAACCGCTTCTATCACAGGTGCAAAACTC
TTTAAGTTAGAAGACGCCAATGGCAAAGATACTGGTTCATTTGCGTTGATTGGTGATGAC
GGTAAACAGTATGCAGCGAATGTTGATCAGAAAACAGGAGCAGTTTCCGTAAACAATG
TCTTACACTGATGCTGACGGTGTCAAACACGACAATGTTAAAGTTGAACTGGGTGGAAGC
GATGGCAAAACCGAAGTTGTAAGTCAACCGATGGCAAAACTTACAGTGTTAGTGATTTA
CAAGGTAAGAGCCTGAAAAGTATTCTATTGCAGCAATTTCTACGCAGAAAACAGAAGAT
CCTTTGGCTGCTATCGATAAAGCACTGTCTCAGGTTGACTCGTTGCGTTCTAACCTAGGT
GCAATTCAAAATCGTTTCGACTCTGCCATCACCAACCTTGGCAACACCGTAAACAACCTG
TCTTCTGCCCCTAGCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCT
CGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCGCAG

Figure 8

31/96

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAACGTCTCTCTTC
TGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGATTGCTAACCGTTT
TACAGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGT
TGCGCAGACCCTGAAGGTGCGCTTTCTGAAATCAACAATAACTTACAGCGTATTCGTGA
ATTGTCAGTACAGGCCACTAATGGTACAACTCTGACTCCGACCTGAATTCAATTCAGGA
TGAAATTACACAACGCCCTTAGTGAAATTGATCGTGTTTCTAACCAGACACAATTTAATGG
TGTAAGGTTCTGGCTTCTGATCAGACTATGAAAATTCAAGTAGGTGCGAACGATGGTGA
AACCATTGAGATTGCCCTTGATAAAATTGATGCTAAAACCTTGGGGCTTGATAACTTTAG
CGTAGCACCAGGAAAAGTTCCAATGTCTCTGCGGTTGCACCTTAAGAGCGAAGCCGCTCC
TGACTTAAGGTAAATGCAACTGATGGTAGTGTGGGAGGTGCTAAAGCATTTCGGTAG
CAATTATAAAAAATGCTGATGTTGAAACTTATTTTGGTACCGGTAATGTACAAGATACAAA
GGATACAAGTATGCGACCGGTACTGCAGGAACAAAAGTTTATCAAGTACAGGTGGAAGG
GCAGACTTATTTGTTGGTCAAGATAATAATACCAACACGAACGGTTTTACATTATTGAA
ACAAAACCTCTACAGGTTATGAAAAAGTTCAGGTGGGTGGTAAGGATGTTTCAGTTAGCAA
CTTTGGTGGTCTGTAACTGCATTTGTTGAAGATAATGGTCTGCCCACATCAGTTGATTT
AGCTGCGGGTAAATGGGTAAAGCATTAGCTTATAATGATGCACCAATGTCTGTTTATTT
TGGGGGAAAAACCTAGATGTCCACCAAGTACAAGATACCCAAGGGAATCCTGTACCTAA
TTCATTTGCTGCTAAAACATCAGACGGCACCTACATTGCAGTAAATGTAGATGCCGCTAC
AGGTAACACGTCTGTTATTACTGATCCTAATGGTAAGGCAGTTGAATGGGCAGTAAAAAA
TGATGGTTCTGCACAGGCAATTATGCGTGAAGATGATAAGGTTTATACAGCCAATATCAC
GAATAAGACGGCAACCAAAGGTGCTGAACTCAGTGCCTCAGATTTGAAAGCCTTAGCAAC
CACAAATCCATTATCCACATTAGACGAAGCTTTGGCAAAAGTTGATAAGTTGCGCAGTTC
TTTGGGTGCAGTACAAAACCGTTTCGACTCTGCCATCACCAACCTTGGCAACACCGTAAA
CAACCTGTCTTCTGCCCCGTAGCCGTATAGAAGATGCTGACTACGCAACCGAAGTGTCTAA
CATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCACAG

Figure 9

32/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAG
CGCCTCTCTTCTGGTCTGCGTATTAACAGCGCTAAAGATGACGCCGCGGGCCAGGCGATT
GCTAACCGCTTTACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGAC
GGTATTTCTCTGGCGCAGACGGCTGAAGGCGCGCTGTCAGAGATTAACAACAACCTTGCAG
CGTATTTCGTGAACTGACCGTTCAGGCCTCTACCGGCACGAACTCTGATTCCGACCTGTCT
TCTATTCAGGACGAAATCAAATCCCGTCTTGATGAAATTGACCGTGTATCTGGTCAGACC
CAGTTCAACGGTGTGAACGTGCTGTCGAAAAACGATTCGATGAAGATTCAGATTGGTGCC
AATGATAACCAGACGATCAGCATTTGGCTTGCAACAAATCGACAGTACCACTTTGAATCTG
AAAGGATTTACCGTGTCCGGCATGGCGGATTTTCAGCGCGGCGAAACTGACGGCTGCTGAT
GGTACAGCAATTGCTGCTGCGGATGTCAAGGATGCTGGGGGTAAACAAGTCAATTTACTG
TCTTACACTGACACCGCGTCTAACAGTACTAAATATGCGGTCTGTTGATTCTGCAACCGGT
AAATACATGGAAGCCACTGTAGTCATTACCGGTACGGCGGCGGCGGTAACGTGTTGGTGCA
GCGGAAGTGGCGGGAGCCGCTACAGCCGATCCGTTAAAAGCACTGGATGCCGCAATCGCT
AAAGTCGACAAATTCCGCTCCTCCCTCGGTGCCGTTCAAACCGTCTGGATTCTGCGGTC
ACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCCAGGACGCC
GACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATTATCCAGCAGGCGGGCAAC TCCGTGCTGTCTAA

Figure 10

33/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTAT
CGAGCGCCTCTCTTCTGGTCTGCGTATTAACAGCGCTAAAGATGACGCCGCGGGCCAGGC
GATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAA
CGACGGTATCTCTCTGGCGCAGACCACTGAAGGCGCGCTGTCTGAAATCAACAACAACCTT
GCAGCGTGTGCGTGAGTTGACCGTT CAGGCGACGACCGGGACTAACTCTGATTCTGACCT
GTCTTCTATT CAGGACGAAATCAAATCCCGTCTGGATGAAATTGATCGCGTTTCCGGTCA
GACCCAGTTCAACGGCGTGAATGTGCTGGCGAAAGATGGTTTCGATGAAGATT CAGGTTGG
CGCGAATGATGGGCAGACTATTAGCATTGATTTGCAGAAGATTGACTCTTCTACATTAGG
ACTGAACGGTTTCTCCGTTTCCGGT CAGTCACTTAACGTTAGTGATTCCATTACTCAAAT
TACCGGTGCCGCCGGGACAAAACCTGTTGGTGTGATTTCACTGCTGTTGCGAAAGATCT
GACTACTGCGACAGGTAAACAGTCGATGTTTCTAGCCTGACGTTACACAACACTCTGGA
TGCGAAAGGGGCTGCTACATCACAGTTCGTCTGTTCAATCCGGCAATGATTTCTACTCCGC
GTCGATTAATCATACAGACGGCAAAGTCACGTTGAATAAAGCCGATGTCGAATACACAGA
CACCGATAATGGACTAACGACTGCGGCTACTCAGAAAGATCAACTGATTAAAGTTGCCGC
TGACTCTGACGGCTCGGCTGCGGGATATGTAACATTCCAAGGTAAAACTACGCTACAAC
GGTTTCAACGGCACTTGATGATAATACTGCGGCAAAGCAACAGATAATAAAGTTGTTGT
TGAATTATCAACAGCAAAACCGACTGCACAGTTCTCAGGGGCTTCTTCTGCTGATCCACT
GGCACTTTTAGACAAAGCTATTGCACAGGTTGATACTTTCCGCTCCTCCCTCGGTGCGGT
GCAAAACCGTCTGGATTCCGCAGTAACCAACCTGAACAACACCACCACCAACCTGTCTGA
AGCGCAGTCCCGTATTCAGGACGCCGACTATGCTACAGAAGTGTCCAACATGTCGAAAGC
GCAGATCATCCAGCAGGCAGGTAACCTCGGTGCTGTCCAA

Figure 11

34/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGC
TGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTC
TGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCTA
ACCGTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTA
TTTCTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTA
TTCGTGAAGTACGGTTTCAAGCTTCTACCGGGACTAACTCTGATTTCGGATCTGGACTCCA
TTCAGGACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTATCCGGTCAGACCCAGT
TCAACGGCGTGAACGTACTGGCAAAAGACGGTTCGATGAAAATTCAGGTTGGTGCGAATG
ACGGCCAGACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAATG
GGTTTAATGTGAACGGCAAAGGGGAAACGGCTAATACGGCAGCAACCCTGAAAGATATGT
CTGGATTACAGCTGCGGCGGCACCGAGGGGAACTGTTGGTGTAACCTCAATATACTGACA
AATCGGCTGTAGCAAGTAGCGTAGATATTCTAAATGCTGTTGCTGGCGCAGATGGAAATA
AAGTTACAACCTAGCGCCGATGTTGGTTTTGGTACACCAGCCGCTGCTGTAACCTATACCT
ACAATAAAGACACTAATTCATATTCCGCCGCTTCTGATGATATTTCCAGCGCTAACCTGG
CTGCTTTCTCAATCCTCAGGCCGGAGATACGACTAAAGCTACAGTTACAATTGGTGGCA
AAGATCAAGATGTAAACATCGATAAATCCGGTAATTTAACTGCTGCTGATGATGGCGCAG
TACTTTATATGGATGCTACCGGTAACCTAACTAAAAATAATGCTGGTGGTGATACACAAG
CTACTTTGGCTAAACTTGCTACTGCTACTGGTGTAAAGCCGCGACCATCCAAACTGATA
AAGGAACATTCACCAGTGACGGTACAGCGTTTGATGGTGCATCAATGTCCATTGATACCA
ATACATTTGCAAATGCAGTAAAAAATGACACTTATACTGCCACTGTAGGTGCTAAGACTT
ATAGCGTAACAACAGGTTCTGCTGCTGCAGACACCGCTTATATGAGCAATGGGGTTCTCA
GTGATACTCCGCCAACTTACTATGCACAAGCTGATGGAAAGTATCACAACTACTGAGGATG
CGGCTGCCGGTAAACTGGTCTACAAAGGTTCCGATGGTAAGTTAACAACGGATACGACTA
GCAAAGCAGAATCAACATCAGATCCGCTGGCAGCTCTTGACGACGCTATCAGCCAGATCG
ACAAATTCCGCTCCTCCCTGGGTGCGGTGCAAAACCGTCTGGATTCCGCAGTGACCAACC
TGAACAACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATG
CGACCGAAGTGTCCAACATGTGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGC
TGGCAAAAGCTAACCAGGTTCCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 12

35/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TACTTCTAA CATTAAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACCGAAGGC
GCGCTGTCTG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
ACCGGAACCTA ACTCTGATTC GGATCTGGAC TCCATTACAGG ACGAAATCAA ATCCCGTCTT
GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
GACGGTTCGA TGAAAATTCA GGTGGGTGCG AATGACGGTG AAATATCAC TATCGACCTG
AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
CGATAAATTA GGGAAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTACT GATTCAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACC ACGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
GAGCGAAGGT AGTGACGGTG CTTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAA
AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGA TTAATTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGCTTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAACCTCCG
GGGTAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTGCA AAGCGCAGAT
TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 13

36 / 96

AACAAATCTCAGTCTTCTCTTAGCTCTGCTA
 TTGAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGG
 CGATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAA
 ATGATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACC
 TGCAGCGTATTTCGTGAACTTTCTGTTGAGGCACTAACGGTACTAACTCTGACAGCGATC
 TTTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGC
 AAATCAGTTTAACGGCGTGAAAGTCCTTGCTGAAAAATAATGAAATGAAAAATTCAGGTTG
 GTGCTAATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACTCTCG
 GCCTGGACGGTTTTAATATCGATGGCGCGCAGAAAGCAACAGGCAGTGACCTGATTTCTA
 AATTTAAAGCGACAGGTAAGTATAATTATGATGTTGGCGGTAAAACTTATACCGTGAATG
 TGGAGAGCGGCGCGGTTAAGAATGATGCTAATAAAGATGTTTTTGTAAGCGCAGCTGATG
 GATCGCTGACGACCAGTAGTGATACTAAAGTATCCGGTGAAAGTATTGATGCAACAGAAAC
 TAGCGAAACTTGCAATAAAATTAGCTGACAAAGGCTCCATTGAATACAAGGGCATTACAT
 TTACTAACAACTGCGCGCAGAGCTTGATGCTAATGGTAAAGGTGTTTTGACCGCAAATA
 TTGATGGTCAAGATGTTCAATTTACTATTGACAGTAATGCACCCACGGGTGCCGGCGCAA
 CAATAACTACAGACACAGCTGTTTACAAAAACAGTGCGGGCCAGTTCACCACTACAAAAG
 TGGAAAATAAAGCCGCAACACTCTCTGATCTGGATCTTAATGCAGCCAAGAAAACAGGTA
 GCACTTTAGTTGTAATGGCGCCACCTACAATGTGACGCGCAGATGGTAAAACGGTAACTG
 ATACTACTCCTGGTGCCCTTAAAGTGATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGA
 TTCTGGTAAACGAAGATGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTA
 TCGACAAGGCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACC
 GTTTCGACTCTGCCATCACCAACCTTGCGCAACACCGTAAACAACCTGTCTTCTGCCCGTA
 GCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCC
 TGCAACAAGCGGGTACCTCTGTTCTGGCGCAG

Figure 14

37/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCG
CTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCATCGAGCGT
CTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCT
AACC GTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGT
ATTTCCGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGT
ATTCTGAACTGACGGTTCAGGCTTCTACCGGGACTAACTCCGATTCTGGATCTGGACTCC
ATT CAGGACGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGCCAGACCCAG
TTCAACGGCGTGAACGTGCTGTCCAAAGATGGCTCGATGAAAATT CAGGTCGGCGCGAAC
GATGGCGAAACGATTACTATTGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCT
GGTTTTAACGTTAACGGTAAAGGTTCTGTAGCGAATACAGCTGCGACAAGCGACGATTTA
AAACTGGCTGGTTTTCTAAGGGCACACAGATACCAATGGCGTGACCGCGTATACAAAC
ACAATTAGTAATGACAAAGCCAAAGCTTCCGATCTGTTAGCTAATATCACCGATGGATCA
GTGATCACTGGGGGAGGGGCAAACGCTTTTGGCGTGGCTGCAAAGAATGGTTACACCTAT
GATGCAGCAAGTAAATCTTATAGTTTTGCTGCAGATGGTGCCGATT CAGCGAAGACGTTA
AGCATCATTAATCCAAACACCGGTGATTCTGTCGAGGCGACAGTGACTATTGGTGGTAAA
GAGCAGAAAGTTAATATTTCCCAGGATGGAAAAATTACTGCGGCAGATGATAATGCGACG
CTGTATTTAGATAAACAGGGAACTTGACAAAAACGAATGCAGGTAACGATACCGCAGCG
ACTTGGGATGGTTTTAATTTCCAACAGCGATTCTACCGGTGCGGTTCCAGTTGGGGTTGCA
ACTACAATTACAATTACTTCTGGTACAGCTTCCGGAATGTCTGTT CAGTCCGCAGGAGCA
GGAATTCAGACCTCAACAAATTCTCAGATTCTTG CAGGTGGTGCATTTGCGGCTAAGGTA
AGTATTGAGGGAGGCGCTGCTACAGACATTTTGGTAGCAAGTAATGGAACATAACAGCG
GCTGATGGTAGTGCACTTTATCTTGATGCGACTACTGGTGGATTCACTACAACGGCTGGA
GGAAATACAGCTGCTTCGTTAGATAATTTAATTGCTAACAGTAAGGATGCTACCTTAACC
GTAACCTCAGGTACCGGCCAGAACACTGTTTATAGCACAAACAGGAAGTGGCGCTCAGTTC
ACCAGTTTAGCAAAAGTAGACACAGTCAATGTCACCAACGCACATGTCAGTGCCGAAGGT
ATGGCAAATCTGACAAAAAGCAATTTTACCATTGATATGGGCGGTACAGGTACAGTAACT
TACACAGTTTCCAATGGGGATGTGAAAGCTGCTGCAAATGCTGATGTTTATGTCGAAGAT
GGTGCACTTTCAGCCAATGCTACAAAAGATGTAACCTACTTTGAACAAAAAAATGGGGCT
ATTACCAACAGCACCGGTGGTACCATCTATGAAACAGCTGATGGTAAGTTAACAACAGAA
GCTACTACTGCATCCAGTTCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGC
TCCATCGACAAATTCCGCTCCTCCCTCGGTGCGGTGCAAAACCGTCTGGATTCCGCGGTC
ACCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATT CAGGACGCC
GACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATCATCCAGCAGGCCGGTAAC
TCCGTGCTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 15

38/96

ATGGCACAAAGTCATTAATACCAACAGC
CTCTCGCTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATC
GAGCGTCTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCG
ATTGCTAACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAAC
GACGGTATTTCTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTA
CAGCGTGTGCGTGAGCTGACTGTTTACGGCGACCAACCGGTACTAACTCTGAGTCTGACCTG
TCTTCTATCCAGGACGAAATCAAATCTCGCCTGGAAGAGATTGATCGTGTTCAGGTCAG
ACTCAATTTAACGGCGTGAATGTTTTGGCTAAAAGATGGGAAAAATGAACATTCAGGTTGGG
GCAAATGATGGACAGACTATCACTATTGATCTGAAAAAGATCGATTTCATCTACACTAAAC
CTCTCCAGTTTTGATGCTACAACTTGGGCACCAGTGTTAAAGATGGGGCCACCATCAAT
AAGCAAGTGGCAGTAGGTGCTGGCGACTTTAAAGATAAAAGCTTCAGGATCGTTAGGTACC
CTAAATTAGTTGAGAAAGACGGTAAGTACTATGTAAATGACACTAAAAGTAGTAAGTAC
TACGATGCCGAAGTAGATACTAGTAAGGGTAAAATTAAGTTCAACTCTACAAATGAAAGT
GGAAGTACTCCTACTGCAGCGACGGAAGTAACTACTGTTGGCCGCGATGTAAATTTGGAT
GCTTCTGCACTTAAAGCCAACCAATCGCTTGTCGTGTATAAAGATAAAAGCGGCAATGAT
GCTTATATCATTAGACCAAAGATGTAACAATAATCAATCAACTTTCAATGCCGCTAAT
ATCAGTGATGCTGGTGTCTTATCTATTGGTGCATCTACAACCGCGCCAAGCAATTTAACA
GCTAACCCGCTTAAGGCTCTTGATGATGCAATTGCATCTGTTGATAAATTCGCTCTTCT
CTCGGTGCCGTTAGAACCCTCTGGATTCTGCCATTGCCAACCTGAACAACCACTACC
AACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCTGACTATGCGACCGAAGTGTCCAAC
ATGTCGAAAGCGCAGATTATCCAGCAGGCCGGTAAGTCCGTGCTGGCAAAAGCCAACCAG
GTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 16

39/96

AACAAATCTCAGTCTTCTCTGAGCTCCGCCAT
TGAACGTCTCTCTTCTGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTCAGGC
GATTGCTAACCGTTTTACAGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAA
TGATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCT
GCAGCGTGACGTGAACTGACTGTTTACAGGCACTAACGGTACTAACTCTGACAGCGATCT
TTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGATCTGAGCA
AACTCAGTTTAAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGG
TGCTAATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACTCTCGG
CCTGGACGGTTTTAATATCGATGGCGCGCAGAAAGCAACTGGCAGTGACCTGATTTCTAA
ATTTAAAGCGACAGGTACTGATAACTATGATGTTGGCGGTGATGCTTATACTGTAAACGT
AGATAGCGGAGCTGTTAAAGATACTACAGGGAATGATATTTTTGTTAGTGACAGCAGATGG
TTCAGTGACAATAAATCTGACACAAACATAGCTGGTACAGGGATTGATGCTACAGCACT
CGCAGCAGCGGCTAAGAATAAAGCACAGAATGATAAATTCACGTTTAAATGGAGTTGAATT
CACAACAACAACCTGCAGCGGATGGCAATGGGAATGGTGTATATTCTGCAGAAATTGATGG
TAAGTCAGTGACATTTACTGTGACAGATGCTGACAAAAAGCTTCTTTGATTACGAGTGA
GACAGTTTACAAAAATAGCGCTGGCCTTTATACGACAACCAAAGTTGATAACAAGGCTGC
CACACTTTCCGATCTTGATCTCAATGCAGCTAAGAAAACAGGAAGCACGTTAGTTGTTAA
CGGTGCAACTTACGATGTTAGTGACAGATGGTAAAACGATAACGGAGACTGCTTCTGGTAA
CAATAAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGA
TGCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTTGGC
TAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCTAT
CACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCGTAGCCGTATCGAAGATGC
TGACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTAC
CTCTGTTCTGGCGCAG

Figure 17

40/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCA
CTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTT
CTGGCTTGCGTATTAACAGCGCGAAGGATGACGAGCGGGTCAGGCGATTGCTAACCGTT
TCACCTCTAACATTAAAGGCCTGACTCAGGCGGCCCCGTAACGCCAACGACGGTATCTCCG
TTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTG
AACTGACGGTTTACGGCTTCTACCGGGACTAACTCCGATTCCGATCTGGAATCCATTACAG
ACGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACG
GCGTGAACGTACTGGCGAAAGACGGTTCAATGAAAATTACAGTTGGTGCGAATGACGGCC
AGACTATCACGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAATGGTTTTTA
ACGTGAATGGTTCCGGTACGATAGCCAATAAAGCGGCGACCATTAGCGACCTGACAGCAG
CGAAAATGGATGCTGCAACTAATACTATAACTACAACAAATAATGCGCTGACTGCATCAA
AGGCGCTTGATCAACTGAAAGATGGTGACACTGTTACTATCAAAGCAGATGCTGCTCAAA
CTGCCACGGTTTTATACATACAATGCATCAGCTGGTAACTTCTCATTCAGTAATGTATCGA
ATAATACTTCAGCAAAAGCAGGTGATGTAGCAGCTAGCCTTCTCCCGCCGGCTGGGCAAA
CTGCTAGTGGTGTTTATAAAGCAGCAAGCGGTGAAGTGAACTTTGATGTTGATGCGAATG
GTAAAATCACAATCGGAGGACAGAAAGCATATTTAACTAGTGATGGTAACTTAACTACAA
ACGATGCTGGTGGTGCGACTGCGGCTACGCTTGATGGTTTATTCAAGAAAGCTGGTGATG
GTCAATCAATCGGGTTTAAAGAACTGCATCAGTCACGATGGGGGGAACAACCTTATAACT
TTAAAACGGGTGCTGATGCTGATGCTGCAACTGCTAACGCAGGGGTATCGTTCACTGATA
CAGCTAGCAAAGAAACCGTTTTTAAATAAAGTGGCTACAGCTAAACAAGGCAAAGCAGTTG
CAGCTGACGGTGATACATCCGCAACAATTACCTATAAATCTGGCGTTCAGACGTATCAGG
CTGTATTTGCCGCAGGTGACGGTACTGCTAGCGCAAAATATGCCGATAAAGCTGACGTTT
CTAATGCAACAGCAACATACACTGATGCTGATGGTGAAATGACTACAATTGGTTCATACA
CCACGAAGTATTCAATCGATGCTAACAAACGGCAAGGTAACCTGTTGATTCTGGAACCTGGTA
CGGGTAAATATGCGCCGAAAGTAGGGGCTGAAGTATATGTTAGTGCTAATGGTACTTTAA
CAACAGATGCAACTAGCGAAGGCACAGTAACAAAAGATCCACTGAAAGCTCTGGATGAAG
CTATCAGCTCCATCGACAAATTCCGTTCTTCCCTGGGTGCTATCCAGAACCGTCTGGATT
CCGCAGTCACCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATT
AGGACGCCGACTATGCGACCGAAGTGTCACATGTGCGAAAGCGCAGATCATTACAGCAGG
CCGGTAACTCCGTGCTGGCAAAAGCCAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGC AGGGTTAA

Figure 18

41/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
ATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
GTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTA
ACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCCGTTGCGCAGA
CCACTGAAGGTGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAGCTGACGG
TTCAGGCTTCTACCGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCA
AGTCTCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACG
TGCTGGCGAAAGACGGTTCGATGAAAATTACAGTTGGTGCGAATGACGGCCAGACTATCA
CGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAGTGGGTTTAATGTGAATG
GTGGCGGGGCTGTTGCTAACACTGCTGCATCTAAAGCTGACTTGGTAGCTGCTAATGCAA
CTGTGGTAGGCAACAAATATACTGTGAGTGCGGGTTACGATGCTGCTAAAGCGTCTGATT
TGCTGGCTGGAGTTAGTGATGGTGATACTGTTTCAAGCAACCATTAAATAACGGCTTCGGAA
CGGCGGCTAGTGCAACGAATTACAAGTATGACAGTGCAAGTAAGTCTTACTCTTTTGATA
CCACAACGGCTTCAGCTGCCGATGTTTCAAGAAATATTTGACCCCGGGCGTTGGTGATACCG
CTAAGGGCACTATTACTATCGATGGTTCTGCACAGGATGTTTCAATCAGCAGTGATGGTA
AAATTACGTCAAGCAATGGAGATAAACTTTACATTGATACAACCTGGGCGCTTAACGAAAA
ACGGCTTTAGTGCTTCTTTGACTGAGGCTAGTCTGTCCACACTTGCAGCCAATAATACCA
AAGCGACAACCATTGACATTGGCGGTACCTCTATCTCCTTTACCGGTAATAGTACTACGC
CGAACACTATTACTTATTTCAGTAACAGGTGCAAAAGTTGATCAGGCAGCTTTTCGATAAAG
CTGTATCAACCTCTGGAAACGATGTTGATTTCACTACCGCAGGTTATAGCGTCGACGGCG
CAACTGGCGCTGTAACAAAAGGTGTTGCTCCGGTTTATATTGATAACAACGGGGCGTTGA
CCACATCTGATACTGTAGATTTTTATCTACAGGATGATGGTTTCACTGACTAACGGCAGCG
GTAAGGCAGTTTATAAAGATGCTGACGGTAAATTGACGACAGATGCTGAAACTAAAGCTG
CAACCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGCTCCATCGACAAATTCC
GCTCCTCCCTCGGTGCGGTGCAGAACCGTCTGGATTCCGCGGTCACCAACCTGAACAACA
CCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCTGACTATGCGACCGAAG
TATCCAACATGTGCGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAG
CTAACCAGGTACCACAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 19

42/96

ATGGCACAAGTCATTAATACCAACAGC
CTCTCGCTGATCACTCAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATC
GAGCGTCTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCG
ATTGCTAACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAAC
GACGGTATTTCTGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTA
CAGCGTGTGCGTGAACGTGACCGTTCAGGCAACCACCGGTACCAACTCCCAGTCTGACCTG
GACTCTATCCAGGACGAAATTAATCCCGTCTGGACGAAATTGATCGCGTATCCGGTCAG
ACCCAGTTCAACGGCGTGAACGTGCTGGCAAAGACGGTTCATGAAAATTCAGGTTGGC
GCCAACGATGGCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACCTTGAAC
CTGACAGGTTTTTAACGTTAACGGTTCGTTCTGTGGCGAATACTGCAGCAACTAAAGCT
GATTTAACCGCTGCTCAACTCTCTGCACCGGGTGCAGCAGACGCAAATGGTACAGTTACT
TATACTGTCAGTGCTGGTTATAAAGAATCCACTGCTGCAGATGTTATTGCTAGCATCAAA
GACGGCAGTGCTCCGACTTCTGCAATTACTGCAACCATTAAATAATGGCTTCGGTGATTCC
AGTGCGCTGACTTCCAATGACTATACTTATGACCCAGCAAAGGCGACTTCACTTACGAC
GTAGCTTCAAGCGCCAATAATACTGCTGCCCAGGTTGAGTCCCTTCCGACGCCGAAAGCA
GGTGATACCGCAAATCTGAAAGTAACCGTTGGTACGACATCGGTTGATGTCGTTCTGGCC
AGTGATGGTAAGATTACAGCAAAGATGGTTCTGCATTATATATCGACAGTACAGGTAAC
CTGACTCAGAACAGTGCTGGCTTGACCTCTGCTAAACTGGCTACTCTGACTGGCCTTCAG
GGCTCTGGTGTGCTTCAACCATCACTACTGAAGATGGCACTAATATTGATATTGCTGCT
AACGGTAATATTGGTCTGACCGGTGTTGCTATCAGTGCTGATTCTCTGCAGTCAGCGACT
AAATCTACGGGCTTTACTGTTGGTACTGGCGCTACAGGTCTGACCGTAGGTACTGATGGT
AAAGTGACTATCGGCGGGACTACTGCTCAGTCCACACCAGCAAAGATGGTTCCCTGACT
ACTGATAACACCACTAAACTGTATCTGCAGAAAGATGGCTCTGTAACCAACGGTTCAGGT
AAAGCGGTCTATGTAGAAGCGGATGGTGATTTCACTACCGACGCTGCAACCAAGCCGCA
ACCACCACCGATCCGCTGAAAGCCCTGGATGAGGCAATCAGCCAGATCGATAAGTTCCGT
TCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACACC
ACTACCAACCTGTCTGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGCGACCGAAGTG
TCCAACATGTCGAAAGCGCAGATCATTCAGCAGGCCGGTAACTCCGTGCTGGCAAAGCC
AACCAGGTACCGCAACAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 20

43/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCAGGTCAGGCGATTGCTAACCGTTT
TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGT
TGCACAGACCCTGAAGGCGCGCTGTCCGAAATCAACAACAACTTACAGCGTGTGCGTGA
ACTGACCGTTTCAAGCAACCACCGGTACCAACTCCCAGTCTGACCTGGACTCTATCCAGGA
CGAAATTAAATCCCGTCTGGACGAAATTGATCGCGTATCCGGTCAGACCCAGTTCAACGG
CGTGAACGTGCTGGCAAAGACGGTTCATGAAAATTACAGGTTGGCGCGAACGATGGCCA
GACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACCTTGAACCTGACAGGTTTAA
CGTTAACGGTTCTGGTTCTGTGGCGAATACTGCAGCAACTAAAGCTGATTTAACCGTGC
TCAACTCTCTGCACCGGTGTCAGCAGACGCAAATGGTACAGTTACTTATACTGTCACTGC
TGGTTATAAAGAATCCACTGCTGCAGATGTTATTGCTAGCATCAAAGACGGCAGTGCTCC
GACTTCTGCAATTACTGCAACCATTAAATAATGGCTTCGGTGATTCCAGTGCGCTGACTTC
CAATGACTATACTTATGACCCAGCAAAAGGCGACTTCACTTACGACGTAGCTTCAAGCGC
CAATAATACTGCTGCCCAGGTTCAAGTCTTCTGACGCCGAAAGCAGGTGATACCGCAA
TCTGAAAGTAACCGTTGGTACGACATCGGTTGATGTCGTTCTGGCCAGTGATGGTAAGAT
TACAGCAAAAAGATGGTTCTGCATTATATATCGACAGTACAGGTAACCTGACTCAGAACAG
TGCTGGCTTGACCTCTGCTAAACTGGCTACTCTGACTGGCCTTCAGGGCTCTGGTGTTC
TTCAACCATCACTACTGAAGATGGCACTAATATTGATATTGCTGCTAACGGTAATATTGG
TCTGACCGGTGTTTCGTATCAGTGCTGATTCTCTGCAGTCAGCGACTAAATCTACGGGCTT
TACTGTTGGTACTGGCGCTACAGGTCTGACCGTAGGTACTGATGGTAAAGTGAATATCGG
CGGGACTACTGCTCAGTCCTACACCAGCAAAAGATGGTTCCTGACTACTGATAACACCAC
TAAACTGTATCTGCAGAAAGATGGCTCTGTAACCAACGGTTCAGGTAAAGCGGTCTATGT
AGAAGCGGATGGTGATTTCACTACCGACGCTGCAACCAAAGCCGCAACCACCACCGATCC
GCTGAAAGCCCTGGATGAGGCAATCAGCCAGATCGATAAGTTCCGTTTCATCCCTGGGTGC
TATCCAGAACCCTCTGGATTCCGCGGTCAACCAACCTGAACAACACCACTACCAACCTGTC
TGAAGCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCACCAACATGTCGAA
AGCGCAGATCATTACAGAGGCCGGTAACCTCCGTGCTGGCAAAGCCAACCAGGTACCGCA
ACAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 21

44/96

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTCTGCGTATTAACAGCGCTAAA
GATGACGCTGCGGGCCAGGCGATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACT
CAGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCGCAGACGGCTGAAGGCGCGCTG
TCAGAGATTAACAACAACCTTGCAGCGTATTCGTGAACTGACCGTTCAGGCCTCTACCGGC
ACGAACTCTGATTCCGACCTGTCTTCTATTTCAGGACGAAATCAAATCCCGTCTTGATGAA
ATTGACCGTGTATCTGGTCAGACCCAGTTCAACGGTGTGAACGTGCTGTCGAAAAACGAT
TCGATGAAGATTCAGATTGGTGCCAATGATAACCAGACGATCAGCATTGGCTTGCAACAA
ATCGACAGTACCACTTTGAATCTGAAAGGATTTACCGTGTCCGGCATGGCGGATTTTCAGC
GCGGCGAAACTGACGGCTGCTGATGGTACAGCAATTGCTGCTGCGGATGTCAAGGATGCT
GGGGGTAAACAAGTCAATTTACTGTCTTACACTGACACCGCGTCTAACAGTACTAAATAT
GCGGTCGTTGATTCTGCAACCGGTAAATACATGGCAGCCACTGTAGTCATTACCAGTACG
GCGGCGGCGGTAACCTGTTGGTGCAACGGAAGTGGCGGGAGCCGCTACAGCCGAACCGTTA
AAAGCACTGGATGCCGCAATCGCTAAAGTCGACAAATTCCGCTCCTCCCTCGGTGCCGTT
CAAAACCGTCTGGATTCTGCGGTCACCAACCTGAACAACACCACCACCAACCTGTCTGAA
GCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCG
CAGATTATCCAGCAGGCG

Figure 22

45/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
ATATCAACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
GTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTACTTCTA
ATATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTGTTGCACAGA
CCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAACTGACGG
TTCAGGCCACTACAGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCA
AATCTCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACG
TGCTGTCCAAAGATGGTTCAATGAAAATTGAGTTCGGCGCAAATGATGGTGAAACCATCA
CGATTGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCTGGTTTTAACGTGAATG
GCGAAGGTGAAACAGCCAATACTGCTGCAACACTTAAAGATATGGTTGGTTTTAAACTCG
ATAATACGGGGGTCACTACAGCTGGAGTTAATAGATATATTGCTGACAAAGCCGTCGCAA
GTAGCACGGATATTTTTGAATGCGGTAGCTGGTGTGATGGCAGTAAAGTTCCACGGAGG
CAGATGTTGGTTTTGGTGCAGCTGCCCCTGGTACGCCAGTGGAATATACTTATCATAAAG
ATACTAACACATATACGGCTTCTGCTTCAGTTGATGCGACTCAACTGGCGGCATTCTCTGA
ATCCTGAAGCGGGTGGTACCACTGCTGCAACAGTAAGTATTGGCAACGGTACAACAGCTC
AAGAGCAAAAAGTCATTATTGCTAAAGATGGTTCTTTAACTGCTGCTGATGACGGTGCCG
CTCTCTATCTTGATGATACTGGTAACTTAAGTAAACTAACGCAGGCACTGATACTCAAG
CTAACTGTCTGACTTAATGGCAAACAATGCTAATGCCAAAACAGTCATTACAACAGATA
AAGGTACATTTACTGCTAATACGACAAAGTTTGATGGGGTAGATATTTCTGTTGATGCTT
CAACGTTTGCTAACGCCGTTAAAAATGAGACTTACACTGCAACTGTTGGTGTAACTTTAC
CTGCGACATATACAGTCAATAATGGCACTGCTGCATCAGCGTATTTAGTCGATGGAAAAG
TGAGCAAACTCCTGCCGAGTATTTTGCTCAAGCTGATGGCACTATTACTAGTGGTGAAA
ATGCGGCTACCAAGTAAAGCTATCTATGTAAGTGCCAATGGTAACTTAACGACTAATACAA
CTAGTGAATCTGAAGCTACTACCAACCCGCTGGCAGCATTGGATGACGCTATCGCGTCTA
TCGACAAATTCCGTTCTTCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCAGTCACCA
ACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAGTCCCGTATTACAGGACGCCGACT
ATGCGACCGAAGTGTCCAACATGTGAAAGCGCAGATCATTCAGCAGGCCGGTAACTCCG
TGCTGGCAAAAGCCAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 23

46/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAATAATAT
CAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTAT
TAACAGCGCGAAGGATGACGCGCAGGTGAGGCGATTGCTAACCGTTTTACTTCTAACAT
TAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCGCAGACCAC
TGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTATTCGTGAACTGACGGTTCA
GGCGACGACCGGAACCTAACCTCCACCTCTGACCTGGACTCCATCCAGGACGAAATCAAATC
CCGTCTTGACGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAACGGCGTGAACGTGCT
GTCTAAAGATGGCTCGATGAAAATTCAGGTCCGGCGGAACGATGGCGAAACGATTACTAT
TGATCTGAAGAAAATTGACTCTGATACGCTGAATCTGGCTGGTTTTAACGTTAACGGTAA
AGGTTCTGTAGCGAATACCGCTGCGACTACAGATAATCTGACATTGGCTGGTTTTACAGC
GGGTACTAAAGCTGCTGATGGCACCCTAATTATAGCAAAAATGTCCAGTTTGCCGCCGC
GACTGCAAGCAATGTACTGGCTGCTGCTAAAGATGGCGACGAAATTACGTTGCTGGTAA
TAACGGCACAGGTATAGCTGCAACTGGGGGACTTATACTTATCATAAGGACTCTAACTC
ATACAGCTTTAGCGCAACGGCTGCATCTAAAGATTCTCTGTTGAGCACACTGGCACCAAA
CGCTGGCGATACATTTACCGCTAAAGTGACTATTGGTTCTAAATCGCAAGAAGTTAACGT
TAGCAAAGATGGTACGATTACATCCAGCGATGGTAAGGCGCTGTATTTAGATGAGAAGGG
CAACCTGACCCAAACAGGTAGTGGCACAAACAAAGCTGCAACCTGGGATAACCTGATGGC
CAATACAGATACTACAGGCAAAGATGCCTATGGTAACTCTGCGGCAGCAGCTGTTGGGAC
AGTAATCGAAGCAAAAGGAATGACCATCACTTCTGCTGGTGGTAATGCTCAGGTGTTAAA
AGACGCGGCTTATAATGCCGCATATGCGACCTCAATTACTACTGGTACTCCGGGTGATGC
GGGAGCCGCGGGAGCCGCTGCAACTGCGGGTAATGCCGCGGTGGGAGCGCTGGGCGCAAC
GGCAGTTGATAATACCACGGCAGATGTTGCCGATATCTCTATCTCAGCTTCGCAAATGGC
GAGCATCCTTCAGGATAAAGATTTACCTTAAGTGATGGTAGTGATACTTACAACGTGAC
CAGCAATGCTGTCACTATCAATGGCAAAGCAGCAAACATTGATGACAGCGGCGCAATCAC
AGACCAAACAGTAAAGTTGTCAATTATTTGCTCATACTAACGGTAGCGTGACTAACGA
TACAGGCTCCACTATTTATGCGACAGAAGATGGTAGCCTGACCACCGATGCAGCAACCAA
AGCCGAAACCACCGCCGATCCCCTGAAAGCTCTGGACGAAGCCATCAGCTCCATCGACAA
ATTCCGCTCCTCCCTCGGTGCGGTGCAAAACCGTCTGGATTCCGCGGTACCAACCTGAA
CAACACCACCACCAACCTGTCTGAAGCGCAGTCCCCTATTGAGGACGCCGACTATGCGAC
CGAAGTGTCCAACATGTCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGC
AAAAGCTAACCAGGTACCACAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 24

47/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTG
ATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTG
TCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAAC
CGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCGGCCCCGTAACGCCAACGACGGTATT
TCTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTTACAGCGTGTG
CGTGAGCTGACTGTTTCAAGGCGACCAACCGGTACCAACTCCAGTCTGATCTGGACTCTATC
CAGGACGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTC
AACGCGCTGAACGTGCTGGCAAAAGACGGTTCCATGAAAATTCAGGTTGGCGCGAATGAT
GGCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGTTGAAACTGACTGGT
TTTAACGTGAATGGTTCTGGTTCTGTGGCGAATACTGCGGCGACTAAAGCGGATTTGGCT
GCTGCTGCAATTGGTACCCCTGGGGCAGCAGATTCTACAGGTGCCATTGCTTACACAGTA
AGTGCTGGGCTGACTAAACTACAGCCGAGATGTACTGTCTAGCCTCGCTGATGGTACG
ACTATTACGCCACAGGCGTGAAAAATGGCTTTGCTGCAGGAGCCACTTCCAATGCCTAT
AACTTAACAAAGATAATAATACATTTACTTATGACACGACTGCTACGACAGCTGAGCTG
CAGTCTTACCTGACTCCGAAAGCGGGCGACACTGCAACATTCAGTGTTGAAATTGGTGGT
ACTACACAAGACGTGCTGCTGTCCAGTGATGGCAAACTCACTGCTAAGGATGGCTCTAAG
CTTTACATTGATACAACCTGGTAATTTAACTCAGAATGGTGGTAATAACGGTGTGGGAACA
CTCGCGGAAGCGACTCTGAGTGGTTTAGCTCTGAACAAAAATGGTTTAACGGCTGTTAAA
TCCACAATTACTACAGCTGATAACACTTCGATTGTACTGAATGGTTCAAGCGATGGTACT
GGTAATGCTGGTACTGAAGGTACGATTGCTGTTACAGGCGCTGTAATTAGTTCAGCTGCT
CTGCAATCTGCAAGCAAAACGACTGGTTTCACTGTTGGTACAGTAGACACAGCTGGTTAT
ATCTCTGTAGGTACTGATGGGAGTGTTCAAGGCATATGATGCTGCGACTTCTGGCAACAAA
GCTTCTTACACCAACACTGACGGTACACTGACTACTGATAACACCACTAAACTGTATCTG
CAGAAAGATGGCTCTGTAACCAACGGTTCAGGTAAAGCGGTCTATGTAGAAGCGGATGGT
GATTTCACTACCGACGCTGCAACCAAGCCGCAACCACCACCGATCCGCTGGCCGCTCTG
GATGACGCAATCAGCCAGATCGACAAGTTCCGTTTCATCCTTGGGTGCTATCCAGAACCGT
CTGGATTCTGCAGTCACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCC
CGTATTCAGGACGCCGACTATGCGACCGAAGTGTCCAATATGTCGAAAGCGCAGATCATC
CAGCAGGCCGGTAACCTCCGTGCTGGCAAAAGCCAACCAGGTACCGCAGCAGGTTCTGTCT
CTGCTGCAGGGTTAA

Figure 25

48/96

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAA
CGTCTCTCTTCTGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGATT
GCTAACCGTTTTACAGCAAATATTAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGAT
GGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAG
CGTATTCGTGAACTTTCTGTTCAAGCACTAACGGTACTAAGTCTGACAGCGATCTTTCT
TCTATCCAGGCTGAAATTAACAAGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCT
CAGTTTAACGGCGTGAAAGTCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCT
AATGATGGTGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTG
GACGGTTTTAATATCGATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTTCTAAATTT
AAAGCGACAGGTACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTTAATGTA
GATAGTGGCGTAGTACAGGATAAAGATGGCAACAAGTTTATGTGAGTACTGCGGATGGT
TCACTTACGACCAGCAGTGATACTCAATTCAAGATTGATGCAACTAAGCTTGACGTGGCT
GCTAAAGATTTAGCTCAAGGGAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACC
GGCACTGTGCTATAGATGCCAAAGGTAATGGTAAATTAACCGCCAATGTTGATGGTAAG
GCTGTTGAATTCATATTTTCGGGGAGTACTGATACATCAGGTACTAGTGCAACCGTTGCC
CCTACGACAGCCCTATACAAAAATAGTGCAGGGCAATTGACTGCAACAAAAGTTGAAAAT
AAAGCAGCGACACTATCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTA
GTTGTTAACGGTGCAACTTACGATGTTAGTGCAGATGGTAAAACGATAACGGGAGACTGCT
TCTGGTAACAATAAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTA
AACGAAGATGCAGCAAATCGTTGCAATCTACCACCAACCCGCTCGAACTATCGACAAA
GCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGAC
TCTGCCATCACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATC
GAAGATGCTGACTACGCGACCGAAGTGTCTAATGTCTCGTGCGCAGATCCTGCAACAA
GCGGGTACCTCTGTTCTGGCACAG

Figure 26

49/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
ATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
GTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTTACTTCTA
ACATTAAAGGCCTGACTCAGGCGGCACGTAACGCCAACGACGGTATCTCTCTGGCGCAGA
CCACCGAAGGTGCGCTGTCTGAAATCAACAACAATTACAGCGTGTACGTGAAGTACCG
TTCAGGCAACCACCGGTACTAATCCGACTCCGACCTGGCTTCTATTACAGGACGAAATCA
AATCCCGTCTGGATGAAATTGACCGCGTATCTGGTCAGACTCAGTTCAACGGCGTGAACG
TGCTGGCAAAAGACGGTTCCATGAAAATTACAGGTAGGTGCTAACGACGGCCAGACTATCA
CTATTGACCTGAAAAAATCGACTCTGATACTCTGGGCCTGAATGGTTTTAACGTGAATG
GTTCTGGGACGATTACCAACAAAGCAGCAACTGTCAAGTGTACTCGCGCAGGCGGTA
CATTGGTGAATGGTGCCTATGATATAAAACCCTAACACAGCGCTGACTACAAGTATG
CCTTCGCGAAATTGAATGATGGTGTGTTTACTATCAATAATGGTAAGGATACTGCCT
ATAAATATAATGCTGCTACAGGTGGGTTTACGACGGATGTCTCCATCTCCGGGGATCCTA
CCGCTGCTGACGCTACTGCTAATAAACTGCCCGTGATGCACTTGCGGCGTCTTTACATG
CTGAGCCGGGTAAACTGTTAATGGTTCTTGACTACGAATGATGGTACGGTAAATTTG
ATACCGATGCCGATGGTAAGATTTCTATTGGTGGTGTGCTGCTTATGTAGATGCAGCAG
GCAACCTGACCACTAACGCAGCAGGTATGACGACTCAAGCAACAACCTACCGATTTGGTTA
CTGCTGCTGCATCTGCTACTGGTAAGGGTGGATCCCTGACCTTTGGTGACACGACGTATA
AAATTGGTCAGGGTACGGCTGGGGTTGATCCTGATGACGCTTCAGATGATGTACTGGGCA
CCATTTCTTACTCTAAATCAGTAAGCAAGGATGTTGTTCTTGCTGATACTAAAGCAACTG
GTAACACGACAACAGTTGATTTCAACTCCGGTATCATGACTTCAAAGGTTAGTTTCGATG
CAGGTACATCAACTGATACATTCAAAGATGCAGATGGTGCTATCACCAAACTAAAGAAT
ACACCACTTCTTATGCTGTAAATAAAGATACTGGTGAAGTTACCGTTGCTGATTATGCTG
CGGTAGATAGCGCCGATAAGGCTGTTGATGATACTAAATATAAACCGACTATCGGCGCGA
CAGTTAACCTGAATTCTGCAGGTAAATTGACCACTGATACCACAGTGCAGGCACAGCAA
CCAAAGATCCTCTGGCTGCCCTGGACGCTGCTATCAGCTCCATCGACAAATTCGGTTCAT
CCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACACCACTA
CCAACCTGTCCGAAGCGCAGTCCCGTATTACAGGACGCCGACTATGCGACCGAAGTGTCCA
ACATGTGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCCAACC
AGGTACCGCAGCAGGTTCTGTCTCTGCTACAGGGTTAA

Figure 27

50/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATC
GAGCGCCTTTCTTCTGGTCTGCGTATTAACAGCGCTAAAGATGACGCTGCGGGCCAGGCG
ATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAAC
GACGGTATTTCTCTGGCGCAGACCACTGAAGGCGCGCTGTCTGAGATTAACAACAACCTTG
CAGCGTGTGCGTGAGTTGACTGTACAGGCGACGACCGGGACTAACTCTGATTCTGACCTG
TCTTCTATCCAGGATGAAATCAAATCCCGTTTAAGCGAAATTGACCGTGTATCTGGTCAG
ACTCAGTTTAACGGCGTGAACTGCTGGCTAAGAATGACACCCTGTCTATTCTAGGTAGGT
GCAAATGACGGTCAGACTATCAATATTGACCTGCAGCAAATCGATTCTCATACACTGGGT
CTGGATGGTTTTACGCGTTAAAAATAATGATGCAGTGAAAACCAGTGCTGCCGTGAATACT
CTTGGGGGGGGGGCAGGTTCTGTTGCTGTGCGACTTCGCAACAACAGTTTGACTGCTATC
ACTGGTCTCGGTAGCGGTGCTATCAGCGAAATTGCTAAAGACGATAATGGTGATTACTAC
GCGCATGTACAGGGACTACGGGTAATACTGCTGATGGTTACTATGCTGTGATATCGAC
AAGGCTACCGGTGAGGTGCGCTCTGAAAGATGGTAACGTAGATACACCGACAGGTACGCCA
ACGACGACAAGCACATATGACTTCACAGACGCTGGTCAAACCGTTTCCTTTGGCACTGAT
GCTGCAACAGCCGGTATCAGCACTGGTGCTTCTCTCGTTAACTTCAGGATGAGAAAGGC
AATGATACTGCTACTTATGCAATCAAAGCACAAAGATGGCAGCCTGTATGCCGCCAACGTT
GATGAGGCTACCGGTAAAGTCACTGTCAAACCGCCAGCTATACTGATGCTGACGGCAAA
GCAGTGACCGATGCCGCTGTAAACTGGGTGGTGACAATGGCACAACCGAAATTGTTGTC
GATGCTGCGTCAGGTAAACTTACGATGCTGGTGCACTGCAAAACGTTGATCTCTCCAGT
GCAACCAACACGGTAACCGCAATCCCGAACGGTAAACCACGTCTCCGCTGGCTGCCCTT
GACGACGCAATCAGCCAGATCGACAAATCCGCTCCTCCCTCGGTGCGGTGCAGAACCGT
CTGGATTCCGCGGTACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAGTCC
CGTATTCTAGGACGCTGACTATGCGACCGAAGTATCCAACATGTGAAAGCGCAGATCATC
CAGCAGGCAGGTAACTCCGTGCTGTCCAAA

Figure 28

51/96

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTCTGCGCATTAAACAGCGCTAAAG
ATGACGCTGCGGGCCAAGCGATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTC
AGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCGCAGACCACTGAAGGCGCACTGT
CTGAAATCAACAACAACCTTGCAGCGTGTTCGTGAACTGACCGTTCAGGCCACTACCGGTA
CTAACTCTGATTCTGACCTGTCTTCAATACAGGACGAAATCAAATCCCGTCTCGATGAAA
TTGACCGCGTATCCGGTCAGACTCAGTTCAACGGCGTTAATGTTCTTTCCAAAGATGGTT
CAATGAAAATTCAGGTTGGTGCGAATGATGGTCAAACCTATCTCCATCGATCTGAAGAAAA
TTGATTCTTCAACTTTGGGGCTGAATGGCTTCTCAGTTTCTAAAAACTCTCTTAATGTCA
GCAATGCTATCACATCTATCCCGCAAGCCGCTAGCAATGAACCTGTTGATGTTAACTTCG
GTGATACTGATGAGTCTGCAGCAATCGCAGCCAAATTGGGGGTTTCCGATACGTCAAGCC
TGTCGCTGCACAACATCCTTGATAAAGATGGTAAGGCAACAGCTGATTATGTTGTTCACT
CAGGTAAAGACTTCTATGCTGCTTCTGTTAATGCCGCTTCAGGTAAAGTAACCTTAAACA
CCATTGATGTTACTTATGATGATTATGCGAACGGTGTGACGATGCCAAGCAAACAGGTC
AGCTGATCAAAGTTTTCAGCAGATAAAGACGGCGCAGCTCAAGGTTTGTCACTTCAAG
GCAAAACTATTTCTGCTGGTGATGCGGCAGACATTCTTAAGAATGGAGCAACAGCTCTTA
AGTTAACTGATCTGAATTTAAGTGATGTTACTGATACTAATGGTAAGGTAACCACAACTG
CGACTGAGCAATTTGAAGGTGCTTCAACTGAGGATCCGCTGGCGCTTCTGGATAAAGCTA
TTGCATCAGTCGACAAATTCGGTCTTCTCTAGGTGCCGTGCAGAACCGTCTCGATTCCG
CTATCACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGG
ACGCCGACTATGCGACCGAAGTGTCCAACATGTGCGAAAGCGCAGATCATCCAGCAGGCA

Figure 29

52/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCG
CTGATCACTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGT
CTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCT
AACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGT
ATTTCTGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGT
ATTCGTGAAGTACGGTTTCAGGCCACTACAGGGACTAACTCCGATTCTGACCTGGACTCC
ATCCAGGACGAAATCAAATCTCGTCTGGACGAAATTGACCGCGTATCTGGTCAGACCCAG
TTCAACGGCGTGAACGTGCTGTCTAAAGATGGCTCGATGAAAATTGAGTTCGGCGCGAAC
GATGGCGAAACGATTACTATTGATCTGAAGAAAATTGACTCTGATACGCTAAATCTGGCT
GGTTTTAACGTGAATGGTGGCTCTGTTGATAATGCCAAGGCGACTGGCAAAGATCTT
ACTGATGCTGGTTTTACGGCAAGCGCAGCTGATGCTAATGGCAAAATCACTTATACCAA
GACACCGTTACTAAATTCGACAAAGCGACAGCGGCTGATGTATTGGGCAAAGCGGCTGCT
GGCGATAGCATTACCTATGCGGGCACTGATACTGGCTTAGGAGTCGCTGCTGATGCCTCG
ACTTACACCTACAATGCAGCCAATAAGTCTTACACTTTTGATGCTACTGGTGTGGCAAAG
GCGGATGCTGGAACGGCACTGAAAGGGTACTTAGGCGCATCTAACACCGGTAAAATTAAT
ATCGGTGGTACCGAGCAAGAAGTTAATATTGCCAAGATGGCTCCATCACCGATACCAAT
GGCGATGCGCTGTATCTCGATAGTACCGGCACTTAACCAAAAATACCGCAATTTGGGG
GCTGCTGATAAAGCAACTGTAGATAAACTGTTTGCTGGTGCTCAGGATGCAACGATCACC
TTCGATAGCGGCATGACAGCTAAATTCGATCAAACCTGCTGGTACCGTTGATTTCAAAGGC
GCGTCTATTTCTGCTGATGCAATGGCATCAACCTTAAATAATGGTTCTTATACAGCCAAC
GTAGGTGGTAAGGCTTATGCCGTAACCGCTGGCGCAGTTTACAGACAGGTGGCGCAGATGTG
TATAAAGATACCACTGGCGCACTGACGACTGAAGATGACGAAACCGTTACCGCGACCTAC
TACGGTTTTTGCTGATGGTAAAGTTTCTGACGGTGAAGGTTCTACTGTCTATAAAGCTGCT
GATGGTTCCATCACTAAAGATGCGACTACCAAGTCTGAAGCAACCACTGACCCTCTGAAA
GCCCTTGACGACGCAATCAGCCAGATCGACAAATTCCGCTCCTCCCTCGGTGCCGTTCAA
AACCGTCTGGATTCCGCCGTCACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCG
CAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCGCAG
ATCATTACAGAGCCGGTAACTCCGTGCTGGCAAAAGCCAACCAAGGTACCGCAGCAGGTT
CTGTCTCTGCTGCAGGGTTAA

Figure 30

53/96

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGA
GCGTCTCTCTTCTGGCCTGCGTATTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGAT
TGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGA
TGGTATTTCTGTTGCGCAGACTACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCA
GCGTGACGTGAACTGACTGTTTACGGCAACTAACGGTACTAACTCTGACAGCGATCTTTC
TTCTATTTCAGGCAGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAAC
TCAGTTTAAACGGCGTGAAAGTCCTTGCCGAAAATAATGAAATGAAAATTCAGGTTGGTGC
TAATGATGGGGAAACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCT
GGACGGCTTTAATATCGATGGCGCGCAGAAAGCAACTGGCAGTGACCTGATTTCTAAATT
TAAAGCGACAGGTACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTTAATGT
AGATAGTGGAGCAGTTCAAATGAGGATGGTGACGCAATTTTTGTTAGCGCTACCGATGG
TTCTCTGACTACTAAGAGTGATACAAAAGTCGGTGGTACAGGTATTGATGCGACTGGGCT
TGCAAAAGCCGCGAGTTTCTTTAGCTAAAGATGCCTCAATTAAATACCAAGGTATTACTTT
CACCAACAAAGGCACTGATGCATTTGATGGCAGTGGTAAACGGCACTCTAACCGCTAATAT
TGATGGCAAAGATGTAACCTTTACTATTGATGCGACAGGGAAGGACGCAACATTAAAAAC
GTCTGATCCTGTTTACAAAATAGTGACAGGTCAGTTCACTACAATAAGGTTGAAAACAA
AGCCGCTACAGCATCGGATCTGGACTTAAATAACGCTAAAAAAGTGGGTAGTTCTTTAGT
TGTAATGGCGCTGATTATGAAGTTAGCGCTGATGGTAAGACAGTAACCTGGGCTTGGCAA
AACTATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAAAGAAGATGCAGC
AAAATCGTTGCAATCTACTACCAACCCGCTCGAAACCATCGACAAGGCATTGGCTAAAGT
TGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCTATCACCAA
CCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCGTAGCCGTATCGAAGATGCTGACTA
CGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGT TCTGGCGCAG

Figure 31

54/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAATA
ATATCAACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGC
GTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACCGTTTTACTTCTA
ACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGATGGTATTTCTGTTGCACAGA
CCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGAACTGACGG
TTCAGGCTTCTACCGGGACTAACTCCGATTCTGGATCTGGACTCCATTTCAGGACGAAATCA
AATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACGGCGTGAACG
TACTGGCGAAAGACGGTTCAATGAAAATTTCAGGTTGGTGCGAATGACGGCCAGACTATCA
CGATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAGTGGGTTTAATGTGAATG
GTAGCGGGCTGTGGCTAATACTGCAGCGACTAAATCTGATTGGGCAGCAGCTCAACTCT
TGGCTCCAGGTAAGTCTGATGCTAATGGTACAGTTACCTATACTGTTGGCGCAGGCCTGA
AAACATCTACAGCTGCAGATGTAATTGCGAGTTTGGCTAATAACGCAAAAGTTAATGCCA
CAATTGCAAAATGGTTTTGGATCGCCAACAGCTACAGATTATACATACAACAGCGCTACAG
GCGATTTTACATATAGTGCAACTATTGCAGCTGGTACAAATTCTGGTGATAGTAACAGTG
CTCAGTTACAATCCTTCTTGACACCAAAAGCGGGCGATACTGCTAACTTAAACGTTAAAA
TTGGTTCTACGTCAATTGACGTTGTATTGGCTAGCGACGGTAAAAATTACCGCGAAAGATG
GTTCAGAACTATTTATTGACGTAGATGGTAACCTCACTCAAAACAATGCTGGGACTGTCA
AAGCAGCCACTCTTGATGCACTGACTAAAAACTGGCATAACAACAGGCACACCGAGTGCCG
TATCTACGGTAATTACAACCTGAAGATGAAACAACCTTCACTCTGGCTGGCGGTACTGATG
CTACTACTTCTGGTGCAATCACTGTAGCAAATGCAAGAATGAGTGCTGAGTCTCTTCAAT
CGGCAACTAAGTCCACAGGATTCACAGTTGATGTTGGAGCTACTGGTACCAGCGCAGGCG
ATATTAAAGTTGATAGTAAAGGTATAGTACAACAACACACAGGTACAGGTTTTGAAGACG
CTTACACCAAAGCTGATGGTTCACTGACTACCGATAATACAACCAATCTGTTTTTGCAAA
AAGACGGAAGTGTGACCAATGGTTCAGGTAAAGCAGTCTATGTTTCAGCGGATGGTAATT
TTACTACTGACGCTGAACTAAAGCTGCAACCACCGCCGATCCACTGAAAGCTCTGGACG
AAGCGATCAGCTCCATCGACAAATTCCGTTCTTCCCTCGGTGCGGTGCAAAACCGTCTGG
ATTCCGCAGTCACCAACCTGAACAACACCACTACTAACCTGTCTGAAGCGCAGTCCCGTA
TTCAGGACGCTGACTATGCGACCGAAGTGTCCAATATGTGCAAGCGCAGATCATCCAGC
AGGCCGGTAACCTCCGTGCTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGC TGCAGGGTTAA

Figure 32

55/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTCTT
CTGGTCTGCGCATTAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCT
TCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATCTCTC
TGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACTTGCAGCGTGTTCTGTG
AGCTGACCGTTTCAGGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCAGG
ACGAAATCAAATCCCGTCTCGATGAAATTGACCGGTATCCGGTCAGACTCAGTTCAACG
GCGTGAACGTACTGGCAAAAGATAACACCATGAAGATTGAGTTGGTGCGAACGATGGTC
AGACTATATCCATCGACCTGCAAAAAATCGACTCTTCTACTCTTGGTTTGAACGGTTTCT
CCGTTTCTAAAAATGCTCTCGAACTAGCGAAGCGATCACTCAGTTGCCGAACGGTGCGA
ATGCACCAATCGCTGTGAAGATGGATGCGTCTGTTCTGACCGATCTTAACATTACTGATG
CTTCCGCTGTTTCGCTGCACAACGTAATAAGGTGGTGTGCGAACGTCTACTTATGTTG
TTCAGTATGGCGATAAGAGCTATGCAGCATCTGTTGATGCGGGAGGTACAGTAAACTGA
ATAAAGCCGACGTAACATATAACGACGCGCAAAATGGTGTTACGAATGCCACCCAGATTG
GTAGTCTGGTTCAGGTTGGTGCTGATGCAACAATGATGCAGTTGGTTTTGTTACCGTGC
AGGGGAAAAACTATGTTGCTAATGACTCATTAGTCAATGCTAATGGCGCTGCTGGCGCTG
CAGCAACTAGAGTTACAATTGATGGTGATGGTAGCCTTGGAGCTAACCAGGCTAAAATTG
AACTTAGCCAAAATGGTGCTACTGCTGCAACATCAGAGTTCGCTGGTGCTTCAACCAACG
ATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAATTCCGTTCTTCTTTGG
GGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAACACCACTACCAACC
TGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAGTGTCCAACATGT
CGAAAGCGCAGATCATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAAA

Figure 33

56/96

ATGGCACAAAGTCATTAATAACCAACAGCCTCTCGCTGATCACTCAAATAATATCAACAAGA
ACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAACAGCG
CGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTTTACTTCTAACATTAAAGGCC
TGA CTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTGCACAGACCACTGAAGGCG
CGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAACTGACGGTTGAGGCGACGA
CCGGAACCTAACTCCACCTCTGACCTGGACTCCATTGAGGACGAAATCAAATCCCGTCTTG
ATGAAATTGACCGCGTATCCGGCCAAACCCAGTTCAACGGCGTGAACGTACTGTCAAAAG
ATGGCTCGATGAAAATTGAGGTGCGCGCAAATGATGGTGAAACCATCACGATTGATCTGA
AAAAGATCGACTCTTCTACATTGAAGCTGACCAGCTTCAATGTTAACGGTAAAGGCGCTG
TTGATAATGCTAAAGCCACTGAAGCAGATCTGACCGCTGCGGGCTTCTCCCAAGGTGCAG
TCGTGAGTGGCAACAGCACCTGGACTAAATCTACTGTTACTACCTTTAATGCAGCAACAG
CTACCGACGTGCTGGCAAGCGTTAGCGGCGGCAGCACTATTAGCGGTTATACCGGTACAA
ACAATGGATTAGGCGTAGCGGCTTCTACTGCATATACCTACAACGCAACCAGCAAGTCTT
ATTCATTTGACGCAACCGCACTTACCAATGGCGATGGTACTGGGGCCACCACTAAAGTTG
CTGATGTGCTGAAAGCCTATGCAGCAAACGGTGATAATACGGCTCAGATCTCCATCGGCG
GAAGCGCTCAGGACGTTAAAAATTGCCAGCGATGGCACCCCTGACTGACGTCAATGGTGATG
CTTTATATATTGGTTCTGACGGCAACCTGACTAAAAACCAGGCCGGCGGTCCAGATGCGG
CAACGTTGGACGGTATTTTCAACGGTGCGAATGGTAATGCAGCAGTTGATGCGAAGATTA
CATTCGGCAGCGGATGACCGTTGATTTACCCAGGCTAGCAAAAAAGTGGATATTAAGG
GCGCAACGGTATCCGCCGAAGATATGGACACTGCGTTAACTGGGCAGGCTTATACCGTAG
CTAACGGCGCACAGTCTTTTGACGTTGCCGCTGGTGGGGCAGTAACCGCTACTACAGGTG
GCGCTACCGTAAATATTGGTGCTGATGGTGAACCTGACGACTGCGACCAACAAGACTGTCA
CAGAACTTATCACGAATTTGCTAACGGCAATATTCTGGATGATGACGGCGCGGCTCTGT
ACAAAGCGGCTGACGGTTCTCTGACCACTGAAGCTACTGGTAAATCCGAAGTGACCACGG
ATCCGCTGAAAGCGCTGGACGATGCTATCGCATCCGTAGACAAATCCGCTCCTCCCTCG
GTGCGGTGCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACCAACC
TGTCTGAAGCGCAGTCCCGCATTGAGGACGCCGACTATGCGACCGAAGTGTCCAATATGT
CGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAGCCAACCAGGTAC
CGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 34

57/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGCGTATTAACAGCGCTAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCGTTT
TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGT
TGCGCAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGA
ACTGACGGTTTACGGCTTCTACCGGGACTAACTCCGATTCCGATCTGGACTCCATTACAGGA
CGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCTGGCCAGACCCAGTTCAACGG
CGTGAACGTACTGGCGAAAGACGGTTCAATGAAAATTCAGGTTGGTGCGAATGACGGCCA
GACTATCACTATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAGTGGGTTTAA
TGTGAATGGTGGCGGGGCTGTTGCTAATACTGCAGCGACTAAAGATGATTGGTTCGCTGC
ATCAGTTTTCAGCTGCGGTAGGTAATGAATACACTGTCTCTGCTGGCCTGTGCGAAATCAAC
TGCTGCTGATGTTATTGCTAGTCTCACAGATGGTGCAGACGTAAGTGCAGGCTGGTGTAAG
CAATGGTTTTGCTGCAGGGGCAACTGGAGATGCTTATAAATTCAATCAAGCAAACAACAC
TTTTACTTACAATACCACCTCAACAGCGGCAGAACTCCAATCTTACCTCACGCCTAAGGC
GGGGGATACCGCAACTTTCTCCGTTGAAATTGGTGGCACCAAGCAGGATGTTGTTCTGGC
TAGTGATGGCAAATCACAGCAAAAGACGGGTCTAACTTTATATTGACACCACAGGGAA
TTTAACCCAAAACGGTGGAGGTACTTTAGAAGAAGCTACCCTCAATGGCTTAGCTTTCAA
CCACTCTGGTCCAGCCGCTGCTGTACAATCTACTATTACTACTGCGGATGGAACCTTCAAT
AGTTCTAGCAGGTTCTGGCGACTTTGGAACAACAAAACTGCTGGGGCTATTAATGTCAC
AGGAGCAGTGATCAGTGCTGATGCACCTTTTCCGCCAGTAAAGCGACTGGGTTTACTTC
TGGCACTTATACCGTAGGTACAGATGGAGTTGTTAAATCTGGTGGCAATGACGTTTATAA
CAAAGCTGACGGGACGGGATTAATACTACTGACAATACCACAAAATATTATTTACAAGATGA
CGGGTCTGTAATAATGGTTCTGGTAAAGCTGTGTATGCTGATGCAACAGGAAAACTAAC
TACTGACGCTGAACTAAAGCCGAAACCACCGCCGATCCCCTGAAAGCTCTGGACGAAGC
GATCAGCTCCATCGACAAATTCCGTTCTTCCCTCGGTGCGGTGCAAAACCGTCTGGATTTC
CGCGGTACCAACCTGAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCA
GGACGCCGACTATGCGACCGAAGTGTCACATGTGCGAAAGCGCAGATCATCCAGCAGGC
CGGTAACCTCCGTGCTGGCAAAGCTAACAGGTACCGCAGCAGGTTCTGTCTCTGCTGCA GGGTTAA

Figure 35

58/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCGTTT
TACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCCGT
TGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAATTACAGCGTATCCGTGA
ACTGACGGTTTCAGGCCACTACCGGTACTAACTCCGATTCTGACCTGGACTCCATCCAGGA
CGAAATCAAATCTCGTCTTGATGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAATGG
CGTGAATGTGTTGTCCAAAGACGGTTCAATGAAAATTGAGGTGGGCGCAAATGATGGTGA
AACCATCACGATTGACCTGAAAAAATCGACTCTTCTACACTGAAGCTGACCAGCTTCAA
CGTCAACGGTAAAGGCGCTGTTGATAATGCAAAGCCACTGAAGCAGATCTGACCGCTGC
GGGCTTCTCCCAAAGTGACAGTTGTGAGTGGCAATAGCACCTGGACTAAATCTACTGTTAC
TACCTTTAATGCAGCAACAGCTACCGATGTGCTGGCTAGCGTTAGTGGCGGCAGCACTAT
TAGCGGTTATGCTGGCACAACAATGGGTTAGGCGTAGCGGCTTCTACTGCATATACCTA
CAACGCAACCAGCAAGTCTTATTCAATTTGACGCAACCGCACTTACTAATGGTGATGGTAC
TGCGGGCTCAACTAAAGTTGCTGATGTTCTGAAAGCCTATGCAGCAAACGGCGATAACAC
GGCTCAGATCTCCATCGGTGGTAGCGCTCAGGAAGTTAAAATTGCCAGCGATGGTACCCT
GACGGATACTAATGGCGATGCTTTATACATTGGTGCTGACGGTAACCTGACGAAAAACCA
GGCCGGCGGCCAGCCGCGGCAACGTTGGACGGTATTTTCAACGGTGCGAATGGTCATGA
TGCAGTTGATGCGAAGATTACCTTCGGCAGCGGCATGACCGTTGACTTCACCCAGGTTAG
CAACAATGTGGATATTAAGGGCGCGACGGTATCCGCCGAAGATATGAACACTGCGTTAAC
CGGTCAGGCTTATACCGTAGCTAACGGCGCACAGTCTTATGACGTTGCCGCTGATGGTGC
AGTAACTGCTACTACAGGTGGAGCGACCGTAAATATTGGTGCTGAGGGTGAAGTACGAC
TGCGGCCAACAAAGACTGTCACAGAACTTATCACGAATTTGCTAACGGCAATATTCTGGA
TGATGACGGCGCGGCTCTGTATAAAGCGGCTGACGGCTCTCTGACCACTGAAGCTACAGG
TAAATCTGAAGCGACCACGGATCCGCTGAAAGCGCTGGACGATGCTATCGCATCCGTAGA
CAAATTCCGTTCTTCCCTGGGTGCCGTGCAGAACCGTCTGGATTCCGCAGTCACCAACCT
GAACAACACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGC
GACCGAAGTGTCCAACATGTGAAAGCGCAGATTATTGAGCAGGCAGGTAAGTCCGTGCT
GGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 36

59/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTAT
CGAGCGCCTCTCTTCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGC
GATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAA
CGACGGTATCTCTCTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTT
GCAGCGTGTGCGTGAGTTGACTGTTTCAGGCGACGACCGGGACTAACTCTGATTCTGACCT
GTCTTCTATTTCAGGACGAAATCAAATCCCGTCTGGATGAAATTGACCGTGTTTCCGGTCA
GACCCAGTTCAACGGCGTGAACGTGCTGGCTAAAAACGGTTCTATGGCGATTTCAGGTTGG
CGCGAATGATGGGCAGACCATCAACATCGACCTGCAGAAAATCGACTCTTCTACTCTGGG
CCTGGGCGGCTTCTCCGTATCTAACAATGCACTGAAACTGAGCGATTCTATCACTCAGGT
TGGTGCGAGTGGTTCACTGGCAGATGTGAACTGAGCTCTGTTGCCTCGGCTCTGGGTGT
AGACGCAAGCACTCTGACTCTGCACAACGTACAGACCCAGCTGGCGCAGCAACAGCTAA
CTATGTTGTCTCTTCTGGTTCTGACAACTACTCAGTATCTGTTGAAGATAGCTCCGGTAC
AGTTACGCTGAACACCACTGATATAGGTTATACCGATACCGCTAATGGCGTTACTACCGG
TTCCATGACTGGTAAGTACGTTAAAGTTGGAGCTGATGCATTGGGTGCTGCTGTAGGTTA
TGTCACCGTACAGGGACAAAACCTTCAAAGCTGATGCTGGCGCGCTGGTTAACTCCAAGAA
TGCTGCTGGTAGTCAGAATGTTACTTCTGCAATTGGCGATATTGCTAATAAAGCGAATGC
TAACATTTACACTGGAACCTCTTCTGCAGATCCACTGGCTCTGCTGGACAAAGCTATCGC
ATCTGTTGATAAATTCGGTCTTCTCTAGGGGCGGTGCAGAACCGTCTGAGCTCTGCTGT
AACCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGC
CGACTATGCGACCGAAGTGTCCAACATGTCGAAAGCGCAGATCATCCAGCAGGCGGGTAA
CTCCGTGCTGTCTAAA

Figure 37

60/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCA
CTCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTT
CTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCCGGTCAGGCGATTGCTAACCGTT
TTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTG
TTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTG
AACTGACGGTTCAGGCTTCTACCGGGACTAACTCTGATTCCGGATCTGGACTCCATTCCAGG
ACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCAACG
GCGTGAACGTACTGGCAAAAGACGGTTCGATGAAAATTGAGTTGGTGCGAACGACGGCC
AGACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAGTGGGTTTA
ACGTAAATGGTAGCGCAGATAAGGCAAGTGTGCGGGCGACAGCTGACGGAATGGTTAAAG
ACGGATATATCAAAGGGTTAACTTCATCTGACGGCAGCACTGCATATACTAAAACCTACAG
CAAATACTGCAGCAAAAGGATCTGATATTCTTGCGGGCGCTTAAGACTGGCGATAAAATTA
CCGCAACAGGTGCAAAATAGCCTTGCTGATAATGCGACATCGACAACCTTATACTTATAATG
CAACCAGCAATACCTTCTCCTATACGGCTGACGGTGTAACCAAACGAATGCTGCAGCAA
ATCTCATACCTGCAGCAGGGAAAACGACAGCTGCATCAGTTACTATTGGTGGGACAGCAC
AGAATGTAAATATTGATGATTTCGGGCAATATTACTTCAAGTGATGGCGATCAACTTTATC
TGGATTCAACAGGTAACCTGACTAAAAACCAGGCCGGCAACCCGAAAAAAGCAACCGTTT
CTGGGCTTCTCGGAAATACGGATGCGAAAGGTACTGCTGTTAAAACAACCATCAAGACAG
AGGCTGGTGTAACAGTTACAGCTGAAGGTAATACAGGTACTGTAAAAATTGAAGGTGCTA
CTGTTTCAGCATCTGCATTTACGGGCATTGCATATTCGCGCAACACCGGTGGGAATACTT
ATGCTGTTGCCGCAAATAATACTACAAATGGTTTCTGCGGGGGGATGACTTAACCCAGG
ATGCTCAAACCTGTTTCAACCTACTACTCGCAAGCCGATGGCACGGTCACGAATAGCGCAG
GCAAAGAAATCTATAAAGACGCTGATGGTGTCTACAGCACAGAGAATAAAACATCGAAGA
CGTCCGATCCATTGGCTGCGCTTGACGACGCAATCAGCTCCATCGACAAATTCCGTTCAT
CCTTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTACCAACCTGAACAACACCACTA
CCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAGTGTCCA
ACATGTCGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCTAACC
AGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGCTAA

Figure 38

61/96

AACAAATCTCAGTCTTCTCTGAGCTCCGCCATTGAACGTCTCTCTTCTGGCCTGCGTA
TTAACAGTGCTAAAGATGACGCAGCAGGTCAGGCGATTGCTAACCGTTTTACAGCAAATA
TTAAAGGTCTGACTCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCGCAGACCA
CTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTGACGTGAACTGACTGTTTC
AGGCAACTAACGGTACTAACTCTGACAGCGATCTTTCTTCTATCCAGGCTGAAATTACTC
AACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCTCAGTTTAAACGGCGTGAAAGTCC
TTGCTGAAAATAATGAAATGAAAATTGAGGTTGGTGCTAATGATGGTGAAACCATCACTA
TCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAATATCGATGGCG
CGCAGAAAGCAACTGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGTAAGTAACT
ATGATGTTGGCGGTGATGCTTATACTGTTAACGTAGATAGCGGAGCTGGGTAATGACTCC
AACTTATTGATAGTGTTTTATGTTTCAAGATAATGCCCCGATGACTTTGTTCATGCAGCTCCAC
CGATTTTGAGAACGACAGCGACTTCCGTCCCAGCCGTGCCAGGTGCTGCCTCAGATTTCAG
GTTATGCCGCTCAATTGCTGCGTATATCGCTTGCTGATTACGTGCAGCTTTCCCTTCAG
GCGGGATTTCATACAGCGGCCAGCCATCCGTTCATCCATATCACCACGTCAAAGGGTGACAG
CAGGCTCATAAGACGCCCCAGCGTCGCCATAGTGCGTTACCGAATACGTGCGCAACAAC
CGTCTTCCGGAGCCTGTCATACGCGTAAACAGCCAGCGCTGGCGCGATTTAGCCCCGAC
ATAGTCCCCTGTTTCGTCCATTTCCGCGCAGACGATGACGTCACTGCCCCGCTGTATGCG
CGAGGTTACCGACTGCGGCCTGAGTTTTTTAAGTGACGTAAAATCGTGTTGAGGCCAACG
CCCATAATGCGGGCAGTTGCCCCGCATCCAACGCCATTCATGGCCATATCAATGATTTTC
TGGTGCGTACCGGGTTGAGAAGCGGTGTAAGTGAACTGCAGTTGCCATGTTTTACGGCAG
TGAGAGCAGAGATAGCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGCACCAACCCCGTCA
GTAGCTGAACAGGAGGGACAGCTGATAGAAACAGAAGCCACTGGAGCACCTCAAAAACAC
CATCATACACTAAATCAGTAAGTTGGCAGCATTACCGCGGAGCTGTTAAAGATACTACAG
GGAATGATATTTTTGTTAGTGACAGCAGATGGTTCACTGACAACTAAATCTGACACAAACA
TAGCTGGTACAGGGATTGATGCTACAGCACTCGCAGCAGCGGCTAAGAATAAAGCACAGA
ATGATAAATTCACGTTTAAATGGAGTTGAATTCACAACAACAACCTGCAGCGGATGGCAATG
GGAATGGTGTATATTCTGCAGAAATTGATGGTAAGTCAGTGACATTTACTGTGACAGATG
CTGACAAAAAAGCTTCTTTGATTACGAGTGAGACAGTTTACAAAAATAGCGCTGGCCTTT
ATACGACAACCAAGTTGATAACAAGGCTGCCACACTTTCCGATCTTGATCTCAATGCAG
CTAAGAAAAACAGGAAGCACGTTAGTTGTTAACGGTGCAACTTACGATGTTAGTGACAGATG
GTAAACGATAACGGGAGACTGCTTCTGGTAACAATAAAGTCATGTATCTGAGCAAATCAG
AAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCAGCAAAATCGTTGCAATCTACCACCA
ACCCGCTCGAAAACCTATCGACAAAGCATTGGCTAAAGTTGACAATCTGCGTTCTGACCTCG
GTGACGTACAAAACCGTTTCGACTCTGCTATCACCACCTTGGCAACACCGTAACAAC
TGTCTTCTGCCCCGTAGCCGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGT
CTCGTGCGCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCGCAG

Figure 39

62/96

AACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGT
CTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACC
GTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTT
CTGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTGTGC
GTGAACTGACCGTTCAGGCAACCACCGGTACCAACTCCCAGTCTGACCTGGACTCTATCC
AGGACGAAATTAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACCCAGTTCA
ACGGCGTGAACGTACTGGCAAAAGACGGTTCATGAAAATTGAGTTGGCGCGAACGATG
GCCAGACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGCTGAACTGACTGGTT
TTAACGTGAATGGCAAAGCAGCGGTTGATAATGCTAAAGCGACGGATGCAAATCTGACTA
CCGCCGGTTTTTACACAAGGCGTTGTGGATTCAAATGGTAATAGTACTTGGACTAAATCAA
CTACGACTAATTTGATGCGGCAACTGCAGTAAACGTACTAGCAGCAGTTAAAGATGGCA
GCACAATCAATTACACCGGTACTGGTAATGGTTTAGGGATTGCTGCAACAAGTGCTTATA
CATATCACGATAGCACTAAATCCTATACCTTTGATTCTACGGGGGCTGCAGTAGCTGGTG
CCGCGTCCAGCCTGCAAGGTACTTTTGGTACAGATACGAATACTGCAAAAATCACCATCG
ATGGTTCTGCTCAAGAAGTAAACATCGCTAAAGATGGGAAAATTACTGATACTGATGGTA
AAGCTTTATATATCGATTCCACTGGTAATTTGACTAAGAACGGCTCTGATACTTTAACTC
AGGCAACATTGAATGATGTCTTACTGGTGCTAATTCAGTTGATGATACAAGGATTGACT
TCGATAGCGGCATGTCTGTCACCCTTGATAAAGTGAACAGCACTGTAGATATCACTGGCG
CATCTATTTAGCCGCTGCAATGACTAATGAGTTGACAGGTAAGGCCTATACCGTAGTAA
ATGGTGCAGAATCTTACGCTGTAGCTACTAATAACACAGTAAAAACGACTGCTGATGCTA
AAAATGTTTATGTTGATGCTAGTGGTAAATTAACCTACTGATGACAAAGCCACTGTTACAG
AACTTATCATGAATTTGCGAATGGCAATATCTATGATGATAAAGGCGCTGCTGTTTATG
CGGCGCGGATGGTTCTCTGACTACAGAACTACAAGTAAATCAGAAGCTACAGCTAACC
CGCTGGCCGCTCTGGACGACGCAATCAGCCAGATCGACAAATTCCGTTTCATCCCTGGGTG
CTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACACCACTACCAATCTGT
CTGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGCGACCGAAGTGTCATATGTCTGA
AAGCGCAGATCATCCAGCAGGCAGGCAACTCCGTGCTGGCAAAA

Figure 40

63/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTC
TTCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCG
CTTCACTTCTAACATCAAAGGTCTGACTCAGGCCGACGTAACGCCAACGACGGTATCTC
TCTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTTGACGCGTGTTCG
TGAAGTACCGGTTACAGGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCA
GGACGAAATCAAATCCCGTCTCGATGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAA
CGGCGTGAACGTACTGGCAAAAGATGGCTCGATGAAAATTGAGGTCGGTGCAAATGATGG
TCAGACAATCAGCATTGATTTGCAGAAGATTGATTCTTCTACTTTAGGGTTAAATGGTTT
TTCTGTTTCCAAAAATGCAGTATCTGTTGGTGATGCTATTACTCAATTGCCTGGCGAGAC
GGCAGCCGATGCACCAGTAACCATCAAGTTTGATGATTGAGTAAAAACTGATTTAAACT
GACCGATGCTTCAGGGTTAAGTCTGCATAACCTCAAAGATGAAAATGGTAATTTAACTAA
CCAGTATGTTGTACAGAATGGCGGAAAAATCTTACGCTGCTACAGTCGCTGCCAATGGTAA
TGTTACGCTGAACAAAGCAAATGTAACCTACAGCGATGTCGCAAACGGTATTGATACCGC
AACGCAGTCAGGCCAGTTAGTTTCAAGTTGGTGAGATTCTACCGGTACGCCAAAAGCATT
CGTGTCTGTCCAAGGTAAAAGCTTTGGCATTGATGACGCCGCTTGAAGAATAACACTGG
TGATGCTACCGCTACTCAACCGGGAACATCTGGGACAACAGTTGTGCGCAGCGTCAATTCA
TCTGAGTACGGGCAAAAATCTGTAGACGCTGATGTAACGGCTTCCACTGAATTCACAGG
TGCTTCAACCAACGATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAATT
CCGTTCTTCTTTGGGGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAA
CACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTGAGGACGCCGACTATGCGACCGA
AGTGTCCAACATGTCGAAAGCGCAGATTATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAA A

Figure 41

64/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTC
TGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCTTCACTT
CTAACATCAAAGGTCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTCTAGCAC
AGACAGCGGAAGGCGCGCTGTCAGAGATTAACAACAACCTGTCAGCGTGTGCGTGAGTTGA
CCGTGACAGGCAACCACTGGTACCAACTCTGATTCCGATCTCTCTTCTATTTCAGGATGAAA
TTAAATCTCGTCTGGATGAAATTGACCGCGTCTCTGGTCAGACCCAGTTTAAACGGCGTGA
ACGTACTGGCTAAAAACGGTTCTATGGCAATTCAGGTTGGCGCGAACGATGGCCAGACTA
TCTCTATCGACCTGCAGAAAATAGACTCTTCTACTCTGGGTCTGAGCGGCTTCTCTGTTT
CTCAGAACTCCCTGAAACTGAGCGATTCTATCACTACGATCGGCAATACTACTGCTGCAT
CGAAGAACGTGGACCTGAGCGCAGTAGCAACTAAACTGGGCGTGAATGCAAGCACCTGA
GCCTGCACGAAGTTCAGGACTCTGCTGGTGACGGTACTGGTACCTTCGTTGTTTCTTCTG
GCAGCGACAACCTATGCTGTGTCTGTAGACGCGGCTCTGGTGACGTTAACCTGAACACCA
CTGACGTCACCTATGATGACGCTACTAATGGTGTTACTGGCGCGACTCAGAACGGTCAGC
TGATCAAAGTAACTTCTGACGCCAACGGTGCAGCTGTTGGTTACGTAACCATTTCAGGGTA
AAAACCTATCAGGCTGGTGCGACCGGTGTTGACGTTCTGGCGAACAGCGGTGTTGCAGCTC
CAACTACAGCTGTTGATACCGGTACTCTGCAACTGAGCGGTACTGGTGCAACTACTGAGC
TGAAAGGTACTGCAACTCAGAACCCACTGGCACTATTGGACAAAGCTATCGCTTCTGTTG
ATAAATTCCGTTCTTCTCTGGGTGCGGTACAGAATCGTCTGAGCTCTGCTGTAACCAACC
TGAATAACACCACCACTAACCTGTCTGAAGCGCAGTCCCGTATTCAGGATGCCGACTATG
CGACCGAAGTGTCAAATATGTCTAAAGCGCAGATCGTTCAGCAGGCCGGTAAC

Figure 42

65/96

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCT
GGTCTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGGCGATTGCTAACCGTTTT
ACGGCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCAAATGATGGTATTTCTGTT
GCGCAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTATTCGTGAA
CTTTCTGTTTCAGGCAACTAACGGTACTAACTCTGACAGCGATCTTTCTTCTATCCAGGCT
GAAATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACCTCAGTTTAAACGGC
GTGAAAGTCCCTTGCTGAAAATAATGAAATGAAAATTCAGGTTGGTGCTAATGATGGTGAA
ACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAAT
ATCGATGGCGCGCAGAAAGCAACAGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGT
ACTGATAATTATGATGTTGGCGGTAAAACCTTATACCGTGAATGTGGAGAGCGGCGCGGTT
AAGAATGATGCTAATAAAGATGTTTTTGTAAAGCGCAGCTGATGGATCGCTGACGACCAGT
AGTGATACTAAAGTATCCGGTGAAAGTATTGATGCAACAGAACTAGCGAAACTTGCAATA
AAATTAGCTGACAAAGGCTCCATTGAATACAAGGGCATTACATTTACTAACAACACTGGC
GCAGAGCTTGATGCTAATGGTAAAGGTGTTTTGACCGCAAATATTGATGGTCAAGATGTT
CAATTTACTATTGACAGTAATGCACCCACGGGTGCCGGCGCAACAATAACTACAGACACA
GCTGTTTACAAAAACAGTGCGGGCCAGTTCACCACTACAAAAGTGGAAAATAAAGCCGCA
ACACTCTCTGATCTGGATCTTAATGCAGCCAAGAAAACAGGTAGCACTTTAGTTGTAAAT
GGCGCCACCTACAATGTCAGCGCAGATGGTAAAACGGTAACTGATACTACTCCTGGTGCC
CCTAAAGTGATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGAT
GCAGCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAGGCATTGGCT
AAAGTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCCATC
ACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCTAGCCGTATCGAAGATGCT
GACTACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGGTACC
TCTGTTCTGGCGCAG

Figure 43

66/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
CAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTCT
GGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACCGTTTC
ACCTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTT
GCACAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGAA
CTGACGGTTTCAGGCTTCTACCGGGACTAACTCTGATTTCGGATCTGGACTCCATTTCAGGAC
GAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGCCAGACCCAGTTCAACGGC
GTGAACGTGCTGGCGAAAGACGGTTCAATGAAAATTCAGGTTGGTGCGAATGACGGCCAG
ACTATCACTATTGATCTGAAGAAAATTGACTCTGATACTCTGGGTTTGAGTGGATTTAAT
GTGAATGGCAAAGGGGCTGTGGCTAACGCAAAAGCGACCGAAGCAGATTTAACGGGGGCT
GGTTTCTCTCAAGGAGCGGTGGATACAAACGGAAATAGTACTTGGACAAAATCAACCACC
ACCAATTACTCAGCTGCAACAACCTGCTGACTTGTATCGACCATTAAGGATGGCTCTACT
GTTACATATGCAGGGACAGACACCGGATTAGGGGTCGCAGCAGCAGGAAATTATACTTAT
GATGCGAACAGTAAATCTTATTCCTTCAATGCCAATGGTCTGACGGGCGCAAATACCGCA
ACTGCACTCAAAGGTTACTTGGGGACAGGTGCTAACACCGCTAAAATTTCTATCGGTGGT
ACAGAGCAGGAAGTGAATATTGCCAAAGATGGCACTATTACAGATACGAATGGTGATGCG
CTCTATCTGGATATTACCGGCAACCTGACTAAGAACTATGCGGGTTTACCACCTGCAGCA
ACGCTGGATAACGTATTAGCTTCCGCAACTGTAAATGCCACTATCAAGTTTGATAGCGGT
ATGACGGTTGATTACACTGCAGGTACTGGCGCGAATATTACAGGTGCATCCATTTCTGCA
GATGACATGGCCGCAAACTGAGCGGAAAGGCGTACACTGTTGCCAATGGTGCTGAGTCT
TATGACGTTGCTGCAGTTACGGGGGCTGTAACAACCTACAGCAGGTAATTCACCTGTGTAT
GCCGATGCAGACGGTAAATTAACGACGAGTGCCAGTAATACGGTTACTCAGACTTATCAC
GAGTTTGCTAATGGTAACATTTATGATGACAAAGGCTCGTCACTGTATAAAGCTGCAGAT
GGCTCTCTGACTTCTGAAGCTAAAGGGAAATCTGAAGCAACCGCCGATCCCCTGAAAGCT
CTGGACGAAGCCATCAGCTCCATCGACAAATTCCGCTCCTCCCTCGGTGCCGTTCAAAC
CGTCTGGATTCTGCGGTGACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAG
TCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCCAATATGTCGAAAGCGCAGATC
ATCCAGCAGGCCGGTAACTCCGTGTTGGCAAAAGCTAACCAGGTACCGCAGCAGGTTCTG
TCTCTGCTGCAGGGTTAA

Figure 44

67/96

GCGCTGTCGACTTCTATCGAGCGCCTCTCTTCTGGTTTGGCGATTAAACAGCGCTA
AAGATGACGCTGCGGGCCAGGCGATTGCTAACCGCTTCACTTCTAACATCAAAGGTCTGA
CTCAGGCCGCACGTAACGCCAACGACGGTATCTCTCTGGCGCAGACCACTGAAGGCGCAC
TGTCTGAAATCAACAACAACCTTGACGCGTGTTCTGTGAAGTACCGTTTCAGGCCACTACCG
GTACTAACTCTGATTCTGACCTGTCTTCAATCCAGGACGAAATCAAATCCCGCTTGGCTG
AAATCGATCGTGTCTCTGGTCAGACCCAGTTCAACGGCGTGAACGTGCTGGCTAAAAACG
GTTCTCTGAATATTTCAGGTTGGCGCGAATGATGGGCAGACCATCTCTATCGATTTCGAGA
AAATAGACTCTTCTGCCCTTGTTTAAAGTGGTTTTAGTGTTGCCGGTGGGGCGCTAAAAT
TAAGCGATACAGTGACGCAGGTGCGCGATGGTTTCAGCCGCGCCAGTTAAAGTGGATCTGG
ATGCAGCAGCAACAGATATTGGTACTGCTTTGGGGCAAAAGGTTAATGCAAGTTCTTTAA
CGTTGCACAATATCTTAGACAAAGATGGTGCGGCAACTGAGAACTATGTTGTTAGCTATG
GTAGTGATAATTACGCTGCATCTGTTGCAGATGACGGGACTGTAACCTCTTAATAAAACGG
ATATTACTTATTTCAGGCGGTGATATTACCGGCGCTACCAAAGATGATACGTTGATTAAAG
TTGCTGCTAATTCTGACGGAGAGGCCGTTGGTTTTCGCTACCGTTTCAGGGTAAGAATTATG
AAATTACAGATGGTGTAACCAACAGTCCACTGCTGCACCAACCGATATTGCTCAGACCA
TTGATCTGGATACGGCTGATGAATTTACTGGGGCTTCCACTGCTGATCCACTGGCACTTT
TAGACAAAGCTATTGCACAGGTTGATACTTTCCGCTCCTCCCTCGGTGCCGTTCAAAACC
GTCTGGATTCCGCAGTCACCAACCTGAACAACACTACTACCAACCTGTCTGAAGCGCAGT
CCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCCAATATGTCGAAAGCGCAGATCA TCCAGCAGGCC

Figure 45

68/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
CAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCT
GGCTTGCGTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTT
ACTTCTAATATTAAAGGCCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTCTG
GCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTGCAGCGTGTGCGTGAA
CTGACCGTACAGGCGACAACCGGAACGAACCTCCGAATCTGACCTGTCTCTATCCAGGAC
GAAATCAAATCCCGTCTGGAAGAGATTGACCGCGTATCCGGCCAGACTCAGTTCAACGGC
GTGAATGTGCTGGCAAAAGACGGCACCATGAAAATTGAGGTAGGCGCGAACGATGGTCAG
ACTATCTCTATCGATCTGAAAAAATCGACTCTTCAACCCTGGGCCTGACCGGTTTTGAT
GTTTCGACGAAAGCGAATATTTCTACGACAGCAGTAACGGGGGCGGCAACGACCACCTTAT
GCTGATAGCGCCGTTGCAATTGATATCGGAACGGATATTAGCGGTATTGCTGCTGATGCT
GCGTTAGGAACGATCAATTTGATAATACAACAGGCAAGTACTACGCACAGATTACCAGT
GCGGCCAATCCGGGCCTTGATGGTGCTTATGAAATCCATGTTAATGACGCGGATGGTTCC
TTCACTGTAGCAGCGAGTGATAAACAAGCGGGTGCTGCTCCGGGTACTGCTCTGACAAGC
GGTAAAGTTCAGACTGCAACCACCACGCCAGGTACGGCTGTTGATGTCACTGCGGCTAAA
ACTGCTCTGGCTGCAGCAGGTGCTGACACGAGTGGCCTGAAACTGGTTCAACTGTCCAAC
ACGGATTCCGCAGGTAAAGTGACCAACGTGGGTTACGGCCTGCAGAATGACAGCGGCACT
ATCTTTGCAACCGACTACGATGGCACCCTGTGACCACGCCGGGCGCAGAGACTGTGACT
TACAAAGATGCTTCCGGTAACAGCACCCTGCGGCTGTCACTGGGTGGCTCTGATGGC
AAAACCAATCTGGTTACCGCCGCTGACGGCAAAACGTACGGTGCGACTGCACTGAATGGT
GCTGATCTGTCCGATCCTAATAACACCGTTAAATCTGTTGCAGACAACGCTAAACCGTTG
GCTGCCCTGGATGATGCAATTGCGATGGTGCACAAATTCCGCTCCTCCCTCGGTGCGGTG
CAAAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACCAACCTGTCTGAA
GCGCAGTCCCGTATTGAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTGAAAGCG
CAGATTATCCAGCAGGCAAGTAACTCCGTGCTGTCCAAAGCTAACCAGGTTCCGCAGCAG
GTTCTGTCTCTGCTGCAGGGTTAA

Figure 46

69/96

AACAAAAACCAGTCTGCGCTGTGACTTCTATCGAGCGCCTCTCTTCTGGT
CTGCGTATTAACAGCGCTAAAGATGACGCCGCGGGCCAGGCGATTGCTAACCGCTTTACT
TCTAACATCAAAGGTCTGACTCAGGCCGCACGTAACGCCAACGACGGTATTTCTCTGGCG
CAGACGGCTGAAGGCGCGCTGTGAGAGATTAACAACAACCTGCAGCGTATTCGTGAACTG
ACCGTTCAGGCCTCTACCGGCACGAACTCTGATTCCGACCTGTCTTCTATTCAGGACGAA
ATCAAATCCCGTCTTGATGAAATTGACCGTGTATCTGGTCAGACCCAGTTCAACGGTGTG
AACGTGCTGTGAAAAACGATTCGATGAAGATTCAGATTGGTGCCAATGATAACCAGACG
ATCAGCATTTGGCTTGCAACAAATCGACAGTACCACCTTTGAATCTGAAAGGATTTACCGTG
TCCGGCATGGCGGATTTTCAGCGCGGCGAAACTGACGGCTGCTGATGGTACAGCAATTGCT
GCTGCGGATGTCAAGGATGCTGGGGGTAAACAAGTCAATTTACTGTCTTACACTGACACC
GCGTCTAACAGTACTAAATATGCGGTCGTTGATTCTGCAACCGGTAAATACATGGAAGCC
ACTGTAGCCATTACCGGTACGGCGGCGGCGGTAACTGTTGGTGCAGCGGAAGTGGCGGGA
GCCGCTACAGCCGATCCGTTAAAAGCACTGGATGCCGCAATCGCTAAAGTCGACAAATTC
CGCTCCTCCCTCGGTGCCGTTCAAAACCGTCTGGATTCTGCGGTCACCAACCTGAACAAC
ACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAA
GTGTCCAACATGTGAAAGCGCAGATTATCCAGCAGGCGGTAACCTCCGTGCTGGCAAA

Figure 47

70/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTC
AAAATAATATCAACAAGAACCAGTCTGCGCTGTGCGAGTTCTATCGAGCGTCTGTCTTCTG
GCTTGCGTATTAAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTTA
CCTCTAACATTAAAGGTCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTTG
CACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTATCCGTGAAC
TGACGGTTCAGGCTTCTACCGGGACTAACTCCGATTCCGGATCTGGACTCCATTACAGGACG
AAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAAACCCAGTTCAACGGTG
TGAACGTACTGGCGAAAGACGGTTCGATGAAAATTACAGGTTGGTGCGAATGACGGCCAGA
CTATCACGATTGATCTGAAGAAAATTGACTCAGATACGCTGGGGCTGAATGGTTTCAACG
TTAATGGCAAAGGCACTATTGCGAACAAGCTGCTACAGTCAGCGATCTGACCGCTGCTG
GTGCAACGGGAACAGGTCCTTATGCTGTGACCACAAACAATACAGCACTCAGCGCTAGCG
ATGCACTGTCTCGCCTGAAAACCGGAGATACAGTTACTACTACTGGCTCGAGTGCTGCGA
TCTATACTTATGATGCGGCTAAAGGGAACCTTCACCACTCAAGCAACAGTTGCAGATGGCG
ATGTTGTAACTTTGCGAATACTCTGAAACCAGCGGCTGGCACTACTGCATCAGGTGTTT
ATACTCGTAGTACTGGTGATGTGAAGTTTGATGTAGATGCTAATGGCGATGTGACCATCG
GTGGTAAAGCCGCGTACCTGGACGCCACTGGTAACCTATCTACAAACAACCCCGGCATTG
CATCTTCAGCGAAATTGTCCGATCTGTTTGCTAGCGGTAGTACCTTAGCGACAACCTGGTT
CTATCCAGCTGTCTGGCACAACCTTATAACTTTGGTGACGCGCAACTTCTGGCGTAACCT
ACACCAAACTGTAAAGCGCTGATACTGTACTGAGCACAGTGCAGAGTGCTGCAACGGCTA
ACACAGCAGTTACTGGTGCGACAATTAAGTATAATACAGGTATTCAGTCTGCAACGGCGT
CCTTCGGTGGTGTGAATACTAATGGTGCTGGTAATTCGAATGACACCTATACTGATGCAG
ACAAAGAGCTCACCACAACCGCATCTTACACTATCAACTACAACGTCGATAAGGATACCG
GTACAGTAACTGTAGCTTCAAATGGCGCAGGTGCAACTGGTAAATTTGCAGCTACTGTTG
GGGCACAGGCTTATGTAACTCTACAGGCAAACCTGACCACTGAAACCACCAGTGCAGGCA
CTGCAACCAAAGATCCTCTGGCTGCCCTGGATGAAGCTATCAGCTCCATCGACAAATTCC
GTTTCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTTACCAACCTGAACAACA
CCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAG
TGTCCAACATGTGCGAAAGCGCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAAG
CCAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 48

71/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTGAGGCGATTGCTAACCGTTT
TACTTCTAATATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAATGACGGTATTTCTGT
TGACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACTTACAGCGTGTGCGTGA
ACTGACCGTTGAGGCGACCAACCGGTACCAACTCCCAGTCTGATCTGGACTCTATCCAGGA
CGAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAACGG
CGTGAACGTACTGGCAAAGACGGTTCATGAAAATTGAGTTGGCGCGAATGATGGCCA
GACCATCACTATCGACCTGAAGAAGATTGACTCTTCTACGTTGAACTGACTGGTTTTAA
CGTGAATGGTTCTGGTTCTGTGGCGAATACTGCGGCGACTAAAGACGAACTGGCTGCTGC
TGCTGCGGCGGCGGGTACAACCTCTGCTGTGCGTACTGACGGCGTGACCAAATATACCGT
AGACGCGAGGGCTTAACAAAGCCACAGCAGCAAACGTGTTTGCAAACCTTGAGATGGTGC
TGTTGTTGATGCTAGCATTTCCAACGGTTTTTGGTGCAGCAGCAGCCACAGACTACACCTA
CAATAAAGCTACAAATGATTTCACTTTCAATGCCAGCATTGCTGCTGGTGCTGCGGCCGG
TGATAGTAACAGCGCAGCTCTGCAATCCTTCTGACTCCAAAAGCAGGTGATACAGCTAA
CCTGAGCGTCAAAATCGGTACGACATCTGTTAATGTTGTTCTGGCGAGCGATGGCAAAT
TACAGCGAAAGATGGCTCAGCTCTGTATATCGACTCAACGGGTAACTGACTCAGAACAG
CGCAGGCACTGTAACAGCAGCAACCCTGGATGGACTGACCAAAAACCATGATGCGACAGG
AGCTGTTGGTGTGATATCACGACCGCAGATGGCGCAACTATCTCTCTGGCAGGCTCTGC
TAACGCGGCAACAGGTACTCAATCAGGTGCAATTACACTGAAAAATGTTTCGTATCAGTGC
TGATGCTCTGCAGTCTGCTGCGAAAGGTACTGTTATCAATGTTGATAATGGTGCTGATGA
TATTTCTGTTAGTAAAACCGGGTGTGTTACTACCGGAGGTGCGCCTACTTATACTGATG
CTGATGGTAAATTAACGACAACCAACACCGTTGATTATTTCTGCAAACTGATGGCAGCG
TAACCAATGGTTCTGGTAAAGGGGTTTACACCGATGCAGCTGGTAAATTCACCTACCGACG
CTGCAACCAAAGCCGCAACCACCACCGATCCGCTGAAAGCCCTTGATGACGCAATCAGCC
AGATCGATAAGTTCCGTTTCATCCCTGGGTGCTATCCAGAACCGTCTGGATTCCGCGGTTA
CCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAGTCCCGTATTGAGGACGCCG
ACTATGCGACCGAAGTGTTCAATATGTCGAAAGCGCAGATCATCCAGCAGGCCGGTAACT
CCGTGTTGGCAAAGCTAACAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 49

72/96

AACAAATCTCAGTCTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCTGGT
CTGCGTATTAACAGCGCAAAAGACGATGCAGCAGGTCAGGCGATTGCTAACCGTTTTACG
GCAAATATTAAAGGTCTGACCCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCG
CAGACCACTGAAGGTGCGCTGAATGAAATTAACAACAACCTGCAGCGTATTCGTGAACCT
TCTGTTTCAGGCAACTAACGGTACTAAGCTCTGACAGCGATCTTTCTTCTATCCAGGCTGAA
ATTACTCAACGTCTGGAAGAAATTGACCGTGTATCTGAGCAAACTCAGTTTAACGGCGTG
AAAGTCCTTGCTGAAAATAATGAAATGAAAATTGAGGTTGGTGCTAATGATGGTGAAACC
ATCACTATCAATCTGGCAAAAATTGATGCGAAAACCTCTCGGCCTGGACGGTTTTAATATC
GATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTTCTAAATTTAAAGCGACAGGTACT
GATAATTATCAAATTAACGGTACTGATACTATACTGTTAATGTAGATAGTGAGTAGTA
CAGGATAAAGATGGCAACAAGTTTATGTGAGTGCTGCGGATGGTTCACCTACGACCAGC
AGTGATACTCAATTCAAGATTGATGCAACTAAGCTTGCACTGGCTGCTAAAGATTTAGCT
CAAGGTAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACCGGCACTGGCGCTATA
CCTGCCACAGGTAATGGTGAATTAACCGCCAATGTTGATGGTAAGGCTGTTGAATTCACCT
ATTTGCGGGGAGTGCTGATACATCAGGTACTAGTGCAACCGTTGCCCTACGACAGCCCTA
TACAAAAATAGTGCGAGGGCAATTGACTGCAACAAAAGTTGAAAATAAAGCAGCGACACTA
TCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTAGTTGTTAACGGTGCA
ACTTACGATGTTAGTGAGATGGTAAACGATAACGGAGACTGCTTCTGGTAACAATAAA
GTCATGTATCTGAGCAAATCAGAAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCAGCA
AAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTGGCTAAAGTT
GACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCTGCCATCACCAC
CTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCGTAGCCGTATCGAAGATGCTGACTAC
GCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCTGCAACAAGCGGGTACCTCTGTT CTGGCACAG

Figure 50

73/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCAC
TCAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTC
TGGCTTGCGTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTT
CACCTCTAACATTAAAGGCCTGACTCAGGCGGCCCGTAACGCCAACGACGGTATCTCCGT
TGCGCAGACCACCGAAGGCGCGCTGTCCGAAATCAACAACAACCTTACAGCGTGTGCGTGA
ACTGACGGTACAGGCCACTACCGGTACTAAGTCTGAGTCTGATCTGTCTTCTATCCAGGA
CGAAATTAAATCCCGTCTGGATGAAATTGACCGCGTATCTGGTCAGACCCAGTTCAACGG
CGTGAACGTGCTGGCAAAAAATGGCTCCATGAAAATCCAGGTTGGCGCAAATGATAACCA
GACTATCACTATCGATCTGAAGCAGATTGATGCTAAACTCTTGGCCTTGATGGTTTTAG
CGTTAAAAATAACGATACAGTTACCCTAGTGCTCCAGTAACTGCTTTTGGTGCTACCAC
CACAAACAATATTAACTTACTGGAATTACCTTTTCTACGGAAGCAGCCACTGATACTGG
CGGAACTAACCCAGCTTCAATTGAGGGTGTTTATACTGATAATGGTAATGATTACTATGC
GAAAATCACCGGTGGTGATAACGATGGGAAGTATTACGCAGTAACAGTTGCTAATGATGG
TACAGTGACAATGGCGACTGGAGCAACGGCAAATGCAACTGTAAGTATGCAAAATACTAC
TAAAGCTACAACATCACTTCAGGCGGTACACCTGTTTACAGATTGATAATACTGCAGGTTT
CGCAACTGCCAACCTTGGTGCTGTTAGCTTAGTAAACTGCAGGATTCCAAGGGTAATGA
TACCGATACATATGCGCTTAAAGATACAAATGGCAATCTTTACGCTGCGGATGTGAATGA
AACTACTGGTGCTGTTTCTGTAAAACTATTACCTATACTGACTCTTCCGGTGCCGCCAG
TTCTCCAACCGCGGTCAAACCTGGGCGGAGATGATGGCAAAACAGAAGTGGTCGATATTGA
TGGTAAACATACGATTCTGCCGATTTAAATGGCGGTAATCTGCAAACAGGTTTGACTGC
TGGTGGTGAGGCTCTGACTGCTGTTGCAAATGGTAAACACGGATCCGCTGAAAGCGCT
GGACGATGCTATCGCATCTGTAGACAAATTCCGTTCTTCCCTCGGTGCGGTGCAAAACCG
TCTGGATTCCGCGGTTACCAACCTGAACAACACCACTACCAACCTGTCTGAAGCGCAGTC
CCGTATTACAGGACGCCGACTATGCGACCGAAGTGTCGAATATGTCGAAAGCGCAGATCAT
CCAGCAGGCCGGTAACTCCGTGTTGGCAAAAGCTAACAGGTACCGCAGCAGGTTCTGTC
TCTGCTGCAGGGTTAA

Figure 51

74/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACT
CAAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCGAGCGTCTGTCTTCT
GGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGATTGCTAACCGTTTT
ACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTCTGTT
GCGCAGACCACCGAAGGCGCGCTGTCTGAAATCAACAACAACTTACAGCGTATTCGTGAA
CTGACGGTTTCAAGCTTCTACCGGGACTAAGTCTGATTTCGGATCTGGACTCCATTACAGGAC
GAAATCAAATCCCGTCTGGACGAAATTGACCGCGTATCCGGTCAAACCCAGTTCAACGGT
GTGAACGTACTGGCGAAAGACGGTTCGATGAAAATTACAGTTGGTGCGAATGACGGCCAG
ACTATCACTATTGATCTGAAGAAAATTGACTCTGATACGCTGGGGCTGAATGGTTTTAAC
GTTAACGGCAAAGGTACTATTGCGAACAAGCGGCAACCATTAGTGATCTGGCGGCGACG
GGGGCGAATGTTACTAACTCAAGCAATATTGTTGTACGACAAAGTTCAATGCCTTGGAT
GCAGCGACTGCATTTAGCAAACCTCAAAGATGGTGATTCTGTTGCCGTTGCTGCTCAGAAA
TATACTTATAACGCATCGACCAATGATTTTACGACAGAAAATACAGTAGCGACAGGCACT
GCAACGACAGATCTTGGCGCTACTCTGAAGGCTGCTGCTGGGCAGAGTCAATCAGGTACA
TATACCTTTGCAAATGGTAAAGTTAACTTTGATGTTGATGCAAGCGGTAATATCACTATT
GGCGGCGAAAAGGCTTTCTTGTTGGTGAGCGCTGACTACTAACGATCCCACCGGCTCC
ACTCCAGCAACGATGTCTTCCCTGTTTAAGGCCGCGGATGACAAAGATGCCGCTCAATCC
TCGATTGATTTTGGCGGGAAAAAATACGAATTTGCTGGTGGAATTCTACTAATGGTGGC
GGCGTTAAATTCAAAGACACGGTGTCTTCTGACGCGCTTTTGGCTCAGGTTAAAGCGGAT
AGTACTGCTAATAATGTAAAAATCACCTTTAACAATGGTCCTCTGTCATTCCTGATCG
TTCCAAAATGGTGTATCTGGCTCCGCGGCATCGAATGCAGCCTACATTGATAGCGAAGGC
GAACTGACAACCTACTGAATCCTACAACACAAATTATTCCGTAGACAAAGACACGGGGGCT
GTAAGTGTTACAGGGGGGAGCGGTACGGGTAAATACGCCGCAAACGTGGGTGCTCAGGCT
TATGTAGGTGCAGATGGTAAATTAACCACGAATACTACTAGTACCGGCTCTGCAACCAA
GATCCACTAAATGCGCTGGATGAGGCAATTGCATCCATCGACAAATTCCGTTCTTCCCTG
GGGGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACACCACTACCAAC
CTGTCTGAAGCGCAGTCCCGTATTACAGACGCCGACTATGCGACCGAAGTGTCCAACATG
TCGAAAGCGCAGATCATCCAGCAGGCCGGTAACTCCGTGTTGGCAAAGCTAACCAGGTA
CCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 52

75/96

AACAAGAACCAGTCTGCGCTGTCTGAGTTCTATCGAGCGTCTGTC
TTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGCGGGTCAGGCGATTGCTAACCG
TTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACGACGGTATTTT
TGTTGCGCAGACCACCGAAGGCGCGCTGTCCGAAATTAACAACAACCTACAGCGTGTGCG
TGAGCTGACTGTTTACGGCGACACCCGGTACTAACTCTGAGTCTGACCTGTCTTCTATCCA
GGACGAAATCAAATCTCGCCTGGAAGAGATTGATCGTGTTCAGTCAGACTCAATTTAA
CGGCGTGAATGTTTTGGCTAAAGATGGGAAAATGAACATTCAGGTTGGGGCAAGTGATGG
ACAGACTATCACTATTGATCTGAAAAAGATCGATTTCATCTACACTAAACCTCTCCAGTTT
TGATGCTACAAACTTGGGCACCAAGTGTTAAAGATGGGGCCACCATCAATAAGCAAGTGGC
AGTAGATGCTGGCGACTTTAAAGATAAAAGCTTCAGGATCGTTAGGTACCCTAAAATTAGT
TGAGAAAGACGGTAAGTACTATGTAAATGACACTAAAAGTAGTAAGTACTACGATGCCGA
AGTAGATACTAGTAAGGGTGAAATTAACCTCAACTCTACAAATGAAAGTGGAAGTACTCC
TACTGCAGCGACGGAAGTAACTACTGTTGGCCGCGATGTAAAATTGGATGCTTCTGCACT
TAAAGCCAACCAATCGCTTGTCGTGTATAAAAGATAAAAGCGGCAATGATGCTTATATCAT
TCAGACCAAAGATGTAACAATAATCAATCAACTTTCAATGCCGCTAATATCAGTGATGC
TGGTGTTTTATCTATTGGTGATCTACAACCGCGCCAAGCAATTTAACAGCTGACCCGCT
TAAGGCTCTTGATGATGCAATTGCATCTGTTGATAAATTCGCTCTTCTCTCGGTGCCGT
TCAGAACCGTCTGGATTCTGCCATTGCCAACCTGAACAACCACTACCAACCTGTCTGA
AGCGCAGTCCCGTATTCAGGACGCTGACTATGCGACCGAAGTGTCCAACATGTGAAAGC
GCAGATTATCCAGCAGGCCGGTAACTCCGTGCTGGCAAAA

Figure 53

76/96

ATGGCACAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAAA
TAATATCAACAAGAACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTCTGGCTT
GCGTATTAACAGCGCGAAGGATGACGCAGCGGGTCAGGCGATTGCTAACCGTTTCACCTC
TAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCTAACGATGGTATCTCTCTGGCGCA
GACCACTGAAGGCGCACTGTCTGAGATTAACAACAACCTTACAACGTGTGCGTGAGTTGAC
TGTACAGGCGACCAACCGGTACTAACTCTGATTCTGACCTGGCTTCTATTTCAGGACGAAAT
CAAATCCCCTTTGTCTGAAATTGACCGCGTATCCGGGCAGACCCAGTTCAACGGCGTGAA
CGTATTGTCTAAAGATGGCTCCCTGAAAATTTCAGGTTGGCGCAAATGATGGTCAGACTAT
CTCTATCGACCTGAAGAAAATTGACTCTGATACTCTGGGTTTGAATGGTTTCAACGTTAA
TGGTTCTGGTACCATTGCAAACAAAGCGGCCACAATCAGTGACTTGACTGCTCAGAAAGC
CGTTGACAACGGTAATGGTACTTATAAAGTTACAACCTAGCAACGCTGCACTTACTGCATC
TCAGGCATTAAGTAAGCTGAGTGATGGCGATACTGTAGATATTGCAACCTATGCTGGTGG
TACAAGTTCAACAGTTAGTTATAAATACGACGCAGATGCAGGTAACCTTCAGTTATAACAA
TACTGCAAACAAAACAAGTGCTGCGGCTGGAACCTCTGGCAGATACTCTTCTCCCGGCAGC
TGGCCAGACTAAAACCGGTACTTACAAGGCTGCTACTGGTGATGTAACTTTAATGTTGA
CGCAACTGGTAATCTGACAATTGGCGGACAGCAAGCCTACCTGACTACTGATGGTAACCT
TACAACAAACAACCTCCGGTGGTGCGGCTACTGCAACTCTTAAAGAGCTGTTTACTCTTGC
TGGCGATGGTAAATCTCTGGGGAACGGCGGTACTGCTACCGTTACTCTGGATAATACTAC
GTATAATTTCAAAGCTGCTGCGAACGTTACTGATGGTGCTGGTGTTCATCGCTGCTGCTGG
TGTAACCTTATACAGCCACTGTTTCTAAAGATGTCATTCTGGCACAACCTGCAATCTGCAAG
TCAGGCAGCAGCAACCGCTACCGACGGTGATACTGTGCAACGATCAACTATAAATCTGG
TGTCATGATCGGTTCCGCTACCTTTACCAATGGTAAAGGTACTGCCGATGGTATGACTTC
TGGTACAACCTCCAGTCGTAGCTACAGGTGCTAAAGCTGTATATGTTGATGGCAACAATGA
ACTGACTTCCACTGCATCTTACGATACGACTTACTCTGTCAACGCAGATACAGGCGCAGT
AAAAGTGGTATCAGGTACTGGTACTGGTAAATTTGAAGCTGTTGCTGGTGCGGATGCTTA
TGTAAGCAAAGATGGCAAATTAACGACAGAAACCACCAGTGCAAGGCACTGCAACCAAAGA
TCCTTTGGCTGCCCTGGATGCTGCTATCAGCTCCATCGACAAATTCCGTTCCCTCCCTGGG
TGCTATCCAGAACCGTCTGGATTCCGCAGTCACCAACCTGAACAACCACTACTAACCT
GTCTGAAGCGCAGTCCCGTATTTCAGGACGCCGACTATGCGACCGAAGTGTCATATATGTC
GAAAGCGCAGATCATCCAGCAGGCCGGTAACTCTGTGTTGGCAAAGCTAACCCAGGTACC
GCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 54

77/96

ATGGCACAAGTCATTAATACCAACAGCC
TCTCGCTGATCACTCAAATAATATCAACAAGAACCAGTCTGCGCTGTCGAGTTCTATCG
AGCGTCTGTCTTCTGGCTTGCGTATTAACAGCGCGAAGGATGACGCCGAGGTCAGGCGA
TTGCTAACCGTTTTACTTCTAACATTAAAGGCCTGACTCAGGCTGCACGTAACGCCAACG
ACGGTATTTCTGTTGCACAGACCACTGAAGGCGCGCTGTCCGAAATCAACAACAACCTTAC
AGCGTATTCGTGAACTGACGGTTTACGGCTTCTACCGGGACTAACTCTGATTCGGATCTGG
ACTCCATTTCAGGACGAAATCAAATCCCGTCTCGACGAAATTGACCGCGTTTCCGGTCAGA
CCCAGTTCAACGGCGTGAACGTGCTGGCGAAAGACGGTTTCGATGAAGATTTCAGGTTGGCG
CGAATGACGGGCAGACCATCTCTATCGATTTGCAGAAAATTGATTCTTCAACGCTGGGAT
TGAAAGGTTTTCTCGGTATCAGGGAACGCATTAAAAGTTAGCGATGCGATAACTACAGTTC
CTGGTGCTAATGCTGGCGATGCCCCGGTTACGGTTAAATTTGGTGCGAACGATACCGCTG
CTGCCGCAATGGCTAAAACATTGGGAATAAGTGATACATCAGGCTTGTCCCTACATAACG
TACAAAGCGCGGATGGTAAAGCGACAGGAACCTATGTTGTTCAATCTGGTAATGACTTCT
ATTTCGGCTTCCGTTAATGCTGGTGGCGTTGTTACGCTTAATACCACCAATGTTACTTTCA
CTGATCCTGCGAACGGTGTTACCACAGCAACACAGACAGGTCAGCCTATCAAGGTCACGA
CGAATAGTGCTGGCGCGGCTGTTGGCTATGTTACTATTCAAGGCAAAGATTACCTTGCTG
GTGCAGACGGTAAGGATGCAATTGAAAACGGTGGTGACGCTGCAACAAATGAAGACACAA
AAATCCAACCTTACCGATGAACTCGATGTTGATGGTTCTGTAAAAACAGCGGCAACAGCAA
CATTTTCTGGTACTGCAACCAACGATCCGCTGGCACTTTTAGACAAAGCTATCTCGCAAG
TTGATACTTTCCGCTCCTCCCTCGGTGCCGTACAAAACCGTCTGGATTCTGCGGTCACCA
ACCTGAATAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACT
ATGCGACCGAAGTGTTCCAAACATGTCGAAAGCGCAGATCATCCAGCAGGCGGGTAACTCTG
TGCTGTCTAAAGCTAACCAGGTACCGCAGCAGGTTCTGTCTCTGCTGCAGGGTTAA

Figure 55

78/96

CTTCTCTTAGCTCTGCTATTGAGCGTCTGTCTTCTGGTCTGCGTATTAACAGCGCAAAAG
ACGATGCAGCAGGTGAGGCGATTGCTAACCGTTTTACGGCAAATATTAAAGGTCTGACCC
AGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCGCAGACCACTGAAGGTGCGCTGA
ATGAAATTAACAACAACCTGCAGCGTATTCTGTAACCTTTCTGTTGAGGCACTAACGGTA
CTAACTCTGACAGCGATCTTTCTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAA
TTGACCGTGTATCTGAGCAAACTCAGTTTAAACGGCGTGAAAGTCCTTGCTGAAAATAATG
AAATGAAAATTGAGTTGGTGCTAATGATGGTGAAACCATTGACCTGCCCCACGATTAG
ATACAACACTCAGTTAGTAACGTGCGAATCTTCATTCTCAGAATGACCCTTTCTCCAGCC
CGCTGCAAATTCAGACGGTGTCTGATAATTCAGCGTGGAGTGCGGGCGGCATTGCTTATA
ATCCTGCCGCCAGTCATTAATAATTTTCTGGCATGAACGATATCGCTGAACCAAGTCTC
ATTCAAACATTTCATCGCGAAATCGTCCGTTAAAGCTCTCAATAAATCCGTTCTGCGTTGG
CTTGCCCGGCTGGATTAAGCGCAACTCAACACCATGCTCAAAGGCCCATGATCCAGTGC
ACGGCAAGTGAACCTCGGCCCCCTGGTCAGTTCTTATCGTCGCCGGATAGCCTCGAAACAG
TGCAATGCTGTCCAGAATACGCGTGACCTGAACGCCTGAAATCCCAAAGGCAACAGTGAC
CGTCAGGCATTCTTTGTGAAATCATCGACGCAGGTAAGACACTTGATCCTGCGACCGGT
GGAAAGTGCGTCCATGACGAAATCCATCGACCAGGTGAGATTGGGCGCCGCGGACGGAG
CAGCGGCAGACGTTCTGTTGCCAGCCCTTTACGACGTCTTCTGCGTTTTACGCCAGGCC
ACTGAGGTGATAAAGCCGGTACACGCGCTTATGATTAACATGAAGCCCTTCACGGCGCAG
CAACTGCCAAATACGACGGTAGCCAAAACGCCTGCGCTCCAGTGCCAGCTCAGTGATGCG
CCCTGATAAATGCGCATCAGCAGCCGGACGGTGAGCCTCATAGCGGCAGGTGACAGGGA
TAAACCTGTAAGCCTGCAGGCACGACGTTGCGACAGACCGGTGCGATCACACATCAACAT
CACGGCTTCCCGCTTCTGGTCTGTCTGTCAGTACTTTTCGCCAAGAGCCACCTGAAGCGCC
TCTTTATCCAGCATGGCTTCGGCAAGCAGCTTCTTGAGTCTGGTGTTCTTCTCTCAAGC
GACTTCAGGCGCTTAACTTCAGGCACCTCCATACCGCCATACTTCTTACGCCAGGTGTAA
AACGTGGCATCGGAAATGGCATGCTTGCGGCAGAGTTACGCGGCGGGTACCCAGCTTCG
GCTTCGCGGAGAATACTGATGATCTGTTCTGTCGGAAAAACGCTTCTTTCATGGGGATGTCC
TCATGTGGCTTATGAAGACATTACTAACATCGGGGTGTACTAATCAACGGGGAGCAGGTC
ACCATCACTATCAATCTGGCAAAAATTGATGCGAAAACTCTCGGCCTGGACGGTTTTAAT
ATCGATGGCGCGCAGAAAGCAACCGGCAGTGACCTGATTCTAAATTTAAAGCGACAGGT
ACTGATAATTATCAAATTAACGGTACTGATAACTATACTGTAAATGTAGATAGTGGAGTA
GTACAGGATAAAGATGGCAAACAAGTTTATGTGAGTGCTGCGGATGGTTCACCTACGACC
AGCAGTGATACTCAATTCAAGATTGATGCAACTAAGCTTGAGTGCTGCTAAAGATTTA
GCTCAAGGTAATAAGATTGTCTACGAAGGTATCGAATTTACAAATACCGGCACTGGCGCT
ATACCTGCCACAGGTAATGGTAAATTAACCGCCAATGTTGATGGTAAGGCTGTTGAATTC
ACTATTTCCGGGAGTGCTGATACATCAGGTACTAGTGCAACCGTTGCCCCCTACGACAGCC
CTATACAAAAATAGTGACGGGCAATTGACTGCAACAAAAGTTGAAAATAAAGCAGCGACA
CTATCTGATCTTGATCTGAACGCTGCCAAGAAAACAGGAAGCACGTTAGTTGTTAACGGT
GCAACTTACGATGTTAGTGCGAGATGGTAAAACGATAACGGAGACTGCTTCTGGTAACAAT
AAAGTCATGTATCTGAGCAAATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCA
GCAAAATCGTTGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTGGCTAAA
GTTGACAATCTGCGTTCTGACCTCGGTGACGTACAAAACCGTTTCGACTCTGCCATCACC
AACCTTGGAACACCGTAAACAACCTGTCTTCTGCCCGTAGCCGTATCGAAGATGCTGAC
TACGCGACCGAAGTGTCTAACATGTCTCGTGCGCAGATCCTGCAACAAGCGGTACCTCT
GTTCTGGCACAGGCTAACC

Figure 56

79/96

AACAAAAACCAGTCTGCGCTGTGCGACTTCTATCGAGCGCCTCTCT
TCTGGTCTGCGCATTAAACAGCGCTAAAGATGACGCTGCGGGCCAGGCGATTGCTAACCGC
TTCACCTTCTAACATCAAAGGTCTGACTCAGGCCGACGTAACGCCAACGACGGTATCTCT
CTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTGCAGCGTGTTCTGT
GAACTGACCGTTTCAAGCCACTACCGGTACTAACTCTGATTCTGACCTGTCTTCAATCCAG
GACGAAATCAAATCCCGTCTCGATGAAATTGACCGCGTATCCGGTCAGACTCAGTTCAAC
GGCGTGAACGTAAGTGGCAAAGATGGCTCGATGAAAATTGAGGTCGGTGCAAATGATGGT
CAGACAATCAGCATTGATTTGCAGAAGATTGATTCTTCTACTTTAGGGTTAAATGGTTTT
TCTGTTTTCAAAAATGCAGTATCTGTTGGTGATGCTATTACTCAATTGCCTGGCGAGACG
GCAGCCGATGCACCAGTAACCATCAAGTTTGATGATTGAGTAAAAACTGATTTAAACTG
ACCGATGCTTCAGGGTTAAGTCTGCATAACCTCAAAGATGAAAATGGTAATTTAACTAAC
CAGTATGTTGTACAGAATGGCGGAAAATCTTACGCTGCTACAGTCGCTGCCAATGGTAAT
GTTACGCTGAACAAAGCAAATGTAACCTACAGCGATGTCGCAAACGGTATTGATACCGCA
ACGCAGTCAGGCCAGTTAGTTTCAAGTTGGTGAGATTCTACCGGTACGCCAAAAGCATTC
GTGTCTGTCCAAGGTAAAAGCTTTGGCATTGATGACGCCGCTTGAAGAATAACACTGGT
GATGCTACCGCTACTCCACCGGGAACATCTGGGACAACAGTTGTGCGCAGCGTCAATTCAT
CTGAGTACGGGCAAAAATCTGTAGACGCTGATGTAACGGCTTCCACTGAATTCACAGGT
GCTTCAACCAACGATCCACTGACTCTGCTGGACAAAGCTATCGCATCTGTTGATAAATTC
CGTTCTTCTTTGGGGGCGGTACAGAACCGTCTGAGCTCCGCTGTAACCAACCTGAACAAC
ACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAA
GTGTCCAACATGTGAAAGCGCAGATTATCCAGCAGGCAGGTAACCTCCGTGCTGTCCAAA

Figure 57

80/96

AACAAAAACCAGTCTGCGCTGTCGACTTCTATCGAACGCCTCTCTTCTGG
CCTGCGTATTAACAGTGCGAAAGATGACGCTGCCGGTCAGGCGATAGCTAACCGTTTCAC
CTCTAACATTAAAGGCCTGACTCAGGCTGCGCGTAACGCCAACGACGGTATTTCTCTGGC
GCAGACCACAGAAGGTGCGTTGTCTGAAATCAACAACAACCTTGCAACGTGTGCGTGAGTT
GACCGTTCAGGCGACGACCGGTACTAACTCTGATTCTGACCTGTCATCTATTAGGACGA
AATCAAATCCCGTCTGGATGAGATTGACCGTGTTTCCGGTCAGACCCAGTTCAACGGCGT
GAATGTACTGGCAAAAGACGGTTCGATGAAGATTAGGTTGGCGCGAATGATGGCCAGAC
TATTAGCATTGATTTACAGAAAATTGACTCTTCTACATTAGGGTTGAATGGTTTCTCCGT
TTCTGCTCAATCACTTAACGTTGGTGATTCAATTACTCAAATTACAGGAGCCGCTGGGAC
AAAACCTGTTGGTGTTGATTTCACTGCTGTTGCGAAAGATCTGACTACTGCGACAGGTAA
AACTGTGATGTTTTCCAGCCTGACGTTACACAACACCCTGGATGCGAAAGGGGCTGCCAC
CGCACAGTTCGTGCTTCAATCCGGTAGTGATTTCTACTCCGCGTCCATTGACCATGCAAG
TGGTGAAGTGACGTTGAATAAAGCCGATGTCGAATACAAAGACACCGATAATGGACTAAC
GACTGCAGCTACTCAGAAAGATCAGCTGATTAAAGTTGCCGCTGACTCTGACGGCGCGGC
TGCGGGATATGTAACATTCCAGGGTAAAAACTACGCTACAACGGCTCCAGCGGCGCTTAA
TGATGACACTACGGCAACAGCCACAGCGAACAAGTTGTTGTTGAATTATCTACAGCAAC
TCCGACTGCGCAGTTCTCAGGGGCTTCTTCTGCTGATCCACTGGCACTTTTAGACAAAGC
CATTGCACAGGTTGATACTTTCCGCTCCTCCCTCGGTGCCGTTCAAACCGTCTGGACTC
TGCGGTAACCAACCTGAACAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCA
GGACGCCGACTATGCGACCGAAGTGCTAACATGTCGAAAGCGCAGATCATCCAGCAGGC
GGGTAACCTCTGTGCTGTCTAAA

Figure 58

81/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
 GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTACAGG ACGAAATCAA ATCCCGTCTC
 GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCTG TGAAAATTCA GGTGGGTGCG AATGACGGTG AAATATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTAA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAAAGTGA TCGGTTAGCT GCCACCTTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
 GGGTAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 59

82/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCGAG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAAGGC
CTGACTCAGG CGGCCCGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
ACAGGGACTA ACTCCGATTG TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
GACGGTTCAA TGAAAATTCA GGTGGGTGCG AATGACGGCG AAACCATCAC GATCGACCTG
AAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTGCT GATTCAGCTT CAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTAATTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAACCTCCG
GGGTAAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCGGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
GTCCCGTATT CAGGACGCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
CATTGAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 60

83/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCAGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
GACGGTTCAA TGAAAATTCA GGTGTTGCG AATGACGGCG AAACCATCAC GATCGACCTG
AAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTA ACGCCGCGGC
GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACC ACGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGA TTAATTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAGATT ATGCTCCAGC AATTGGCACT GCTGTAAATG TGAATCCGC
GGGTAAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGTTGCTGC
CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGCGGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
CATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 61

84/96

ATGGCACAAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTAATTCTAA CATTAAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTGAGG ACGAAATCAA ATCCCGTCTC
GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
GACGGTTCTG TGAAAATTCA GGTGGGTGCG AATGACGGTG AACTATCAC TATCGACCTG
AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTACT GATTCAGCTA AAAACGTGA TCGGTTAGCT GCCACCCCTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTAATTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAACCTCCG
GGGTAAAATC ACTACTGAGA CTACCAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 62

85/96

ATGGCACAAAGTCATTAATACCAACAGCCTCTCGCTGATCACTCAAATAATATCAACAAG
AACCAGTCTGCGCTGTGAGTTCTATCGAGCGTCTGTCTTCTGGCTTGCGTATTAACAGC
GCGAAGGATGACGCCGCGAGGTCAGGCGATTGCTAACCGTTTTACTTCTAACATTAAAGGC
CTGACTCAGGCGGCCCGTAACGCCAACGACGGTATTTCTGTTGCGCAGACCACCGAAGGC
GCGCTGTCCGAAATCAACAACAACCTTACAGCGTATTCGTGAACTGACGGTTCAGGCCACT
ACAGGGACTAACTCCGATTCTGACCTGGACTCCATCCAGGACGAAATCAAATCTCGTCTT
GATGAAATTGACCGCGTATCCGGCCAGACCCAGTTCAACGGCGTGAACGTGCTGGCGAAA
GACGGTTCAATGAAAATTGAGGTTGGTGCGAATGACGGCGAAACCATCACGATCGACCTG
AAAAAATCGATTCTGATACTCTGGGTCTGAATGGCTTTAACGTAAATGGTAAAGGTACT
ATTACCAACAAAGCTGCAACGGTAAGTGATTTAACTTCTGCTGGCGCGAAGTTAAACACC
ACGACAGGTCTTTATGATCTGAAAACCGAAAATACCTTGTTAACTACCGATGCTGCATTC
GATAAATTAGGGAATGGCGATAAAGTCACAGTTGGCGGCGTAGATTATACTTACAACGCT
AAATCTGGTGATTTTACTACCACTAAATCTACTGCTGGTACGGGTGTAGACGCCGCGGCG
CAGGCTGCTGATTACGCTTCAAACGTGATGCGTTAGCTGCCACCCTTCATGCTGATGTG
GGTAAATCTGTTAATGGTTCTTACACCACAAAAGATGGTACTGTTTCTTTCGAAACGGAT
TCAGCAGGTAATATCACCATCGGTGGAAGCCAGGCATACGTAGACGATGCAGGCAACTTG
ACGACTAACACGCTGGTAGCGCAGCTAAAGCTGATATGAAAGCGCTGCTCAAAGCAGCG
AGCGAAGGTAGTGACGGTGCCTCTCTGACATTCAATGGCACAGAATATACCATCGCAAAA
GCAACTCCTGCGACAACCACTCCAGTAGCTCCGTTAATCCCTGGTGGGATTACTTATCAG
GCTACAGTGAGTAAAGATGTAGTATTGAGCGAAACCAAAGCGGCTGCCGCGACATCTTCA
ATTACCTTTAATTCCGGTGTACTGAGCAAACTATTGGGTTTACCGCGGGTGAATCCAGT
GATGCTGCGAAGTCTTATGTGGATGATAAAGGTGGTATCACTAACGTTGCCGACTATACA
GTCTCTTACAGCGTTAACAAGGATAACGGCTCTGTGACTGTTGCCGGGTATGCTTCAGCG
ACTGATACCAATAAAGATTATGCTCCAGCAATTGGTACTGCTGTAAATGTGAACTCCGCG
GGTAAATCACTACTGAGACTACCAAGTGCTGGTTCTGCAACGACCAACCCGCTTGCTGCC
CTGGACGACGCAATCAGCTCCATCGACAAATTCCGTTCTTCCCTGGGTGCTATCCAGAAC
CGTCTGGATTCCGCGAGTCACCAACCTGAACAACCACTACCAACCTGTCCGAAGCGCAG
TCCCGTATTACAGGACGCCGACTATGCGACCGAAGTGTCCAACATGTCGAAAGCGCAGATC
ATTACAGCAGGCCGGTAACTCCGTGCTGGCAAAAGCTAACCAAGGTACCGCAGCAGGTTCTG
TCTCTGCTGCAGGGTTAA

Figure 63

86/96

ATGGCACAAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCAGG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
ACAGGGACTA ACTCCGATTC TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
GACGGTTCAA TGAAAATTCA GGTGGGTGCG AATGACGGCG AAACCATCAC GATCGACCTG
AAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAACCTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAACCTACG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGCACT GCTGTAAATG TGAACTCCGC
GGGTAAAATC ACTACTGAGA CTACCAGTGC TGTTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGCGGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
CATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 64

87/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTACTTCTAA CATTAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTACAGG ACGAAATCAA ATCCCGTCTC
GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
GACGGTTCGA TGAAAATTCA GGTGCGTGC AATGACGGTG AACTATCAC TATCGACCTG
AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACACC
ACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAATACCG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTACT GATTCAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTTCTTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
GGGTAAAATC ACTACTGAGA CTACCAAGTGC TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGCAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 65

88/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCAGG TCAGGCGATT GCTAACCGTT TTA CTCTCTAA CATTAAAGGC
CTGACTCAGG CGGCCCCGTAA CGCCAACGAC GGTATTTCTG TTGCGCAGAC CACCGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATTCGTG AACTGACGGT TCAGGCCACT
ACAGGGACTA ACTCCGATTG TGACCTGGAC TCCATCCAGG ACGAAATCAA ATCTCGTCTT
GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT GCTGGCGAAA
GACGGTTCAA TGAAAATTCA GGTTGGTGCG AATGACGGCG AAACCATCAC GATCGACCTG
AAAAAAATCG ATTCTGATAC TCTGGGTCTG AATGGCTTTA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTA ACTTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTA ACTACCG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC AGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACTAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
GCAGGCTGCT GATTCAGCTT CAAAACGTGA TGCGTTAGCT GCCACCCTTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TCAAAGCAGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACAGAATATA CCATCGCAAA
AGCAACTCCT GCGACAACCA CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTACTTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATC ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGA ACTCCGC
GGGTAAAATC ACTACTGAGA CTACCAAGTG TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCAATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CCGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
CATTCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCTAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 66

89/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
 AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
 GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTA CTCTCTAA CATTAAAGGC
 CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACCGAAGGC
 GCGCTGTCTG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
 ACCGGAACTA ACTCTGATTC GGATCTGGAC TCCATTACAG ACGAAATCAA ATCCCGTCTT
 GATGAAATTG ACCGCGTATC CGGCCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
 GACGGTTCGA TGAAAATTCA GGTGTTGTCG AATGACGGTG AACTATCAC TATCGACCTG
 AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTTA ACGTAAATGG TAAAGGTACT
 ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAAC TTCTG CTGGCGCGAA GTTAAACAC
 CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTTG TTAAC TACC GATGCTGCATT
 CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
 TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTGTAG ACGCCGCGGC
 GCAGGCTACT GATTCAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTTC ATGCTGATGT
 GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
 TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
 GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
 GAGCGAAGGT AGTGACGGTG CTTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAA
 AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTA CTTATCA
 GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
 AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
 TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
 AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
 GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGA ACTCCGC
 GGGTAAAATC ACTACTGAGA CTACAGTG C TGTTCTGCA ACGACCAACC CGCTTGCTGC
 CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
 CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
 GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTC AACATGTCGA AAGCGCAGAT
 TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
 GTCTCTGCTG CAGGGTTAA

Figure 67

90/96

ATGGCACAAG TCATTAATAC CAACAGCCTC TCGCTGATCA CTCAAAATAA TATCAACAAG
AACCAGTCTG CGCTGTCGAG TTCTATCGAG CGTCTGTCTT CTGGCTTGCG TATTAACAGC
GCGAAGGATG ACGCCGCGGG TCAGGCGATT GCTAACCGTT TTA CTCTCTAA CATTAAAGGC
CTGACTCAGG CTGCACGTAA CGCCAACGAC GGTATTTCTG TTGCACAGAC CACTGAAGGC
GCGCTGTCCG AAATCAACAA CAACTTACAG CGTATCCGTG AGCTGACGGT TCAGGCTTCT
ACCGGGACTA ACTCTGATTC GGATCTGGAC TCCATTCAGG ACGAAATCAA ATCCCGTCTC
GACGAAATTG ACCGCGTATC CGGTCAGACC CAGTTCAACG GCGTGAACGT ACTGGCAAAA
GACGGTTCGA TGAAAATTCA GGTGCGTGCG AATGACGGTG AAATATCAC TATCGACCTG
AAGAAAATCG ATTCTGATAC TCTGGGTCTG AATGGTTTAA ACGTAAATGG TAAAGGTACT
ATTACCAACA AAGCTGCAAC GGTAAGTGAT TTAATTCTG CTGGCGCGAA GTTAAACAC
CACGACAGGT CTTTATGATC TGAAAACCGA AAATACCTG TTAATACCG ATGCTGCATT
CGATAAATTA GGAATGGCG ATAAAGTCAC CGTTGGCGGC GTAGATTATA CTTACAACGC
TAAATCTGGT GATTTTACTA CCACCAAATC TACTGCTGGT ACGGGTG TAG ACGCCGCGGC
GCAGGCTACT GATTGAGCTA AAAAACGTGA TCGGTTAGCT GCCACCCTC ATGCTGATGT
GGGTAAATCT GTTAATGGTT CTTACACCAC AAAAGATGGT ACTGTTTCTT TCGAAACGGA
TTCAGCAGGT AATATCACCA TCGGTGGAAG CCAGGCATAC GTAGACGATG CAGGCAACTT
GACGACTAAC AACGCTGGTA GCGCAGCTAA AGCTGATATG AAAGCGCTGC TTAAAGCCGC
GAGCGAAGGT AGTGACGGTG CCTCTCTGAC ATTCAATGGC ACTGAATATA CTATCGCAAA
AGCAACTCCT GCGACAACCT CTCCAGTAGC TCCGTTAATC CCTGGTGGGA TTTCTTATCA
GGCTACAGTG AGTAAAGATG TAGTATTGAG CGAAACCAA GCGGCTGCCG CGACATCTTC
AATTACCTTT AATTCCGGTG TACTGAGCAA AACTATTGGG TTTACCGCGG GTGAATCCAG
TGATGCTGCG AAGTCTTATG TGGATGATAA AGGTGGTATT ACTAACGTTG CCGACTATAC
AGTCTCTTAC AGCGTTAACA AGGATAACGG CTCTGTGACT GTTGCCGGGT ATGCTTCAGC
GACTGATACC AATAAAGATT ATGCTCCAGC AATTGGTACT GCTGTAAATG TGAATCCGC
GGGTAAAATC ACTACTGAGA CTACCAAGTG TGGTTCTGCA ACGACCAACC CGCTTGCTGC
CCTGGACGAC GCTATCAGCT CCATCGACAA ATTCCGTTCT TCCCTGGGTG CTATCCAGAA
CCGTCTGGAT TCCGAGTCA CCAACCTGAA CAACACCACT ACCAACCTGT CTGAAGCGCA
GTCCCGTATT CAGGACGCCG ACTATGCGAC CGAAGTGTCC AACATGTCGA AAGCGCAGAT
TATCCAGCAG GCCGGTAACT CCGTGCTGGC AAAAGCCAAC CAGGTACCGC AGCAGGTTCT
GTCTCTGCTG CAGGGTTAA

Figure 68

91/96

ATGCGACGTATAGAACGAATACCGGGGTTATCGGCGTAAGCGGGGCAAA
GTTTACGATTTATTTTTTGGCTTAATGACACGAACAGCAACGAGGAAGGG
GAGTATTTTCGACCGCTAGAAAAAATTCTAAAGGTTGTGAGTGACCAGAC
GATAACAGGGTTGACGGCGACGAAGCCGAAGGGTGGAAGCCCAATACTT
AAACCGTAGACTTGAAAACAGGAAAATGAATCATGGCACAAGTCATTAAT
ACCAACAGCCTCTCGCTGATCACTCAAAATAATATCAACAAGAACCAGTC
TGCGCTGTGCGACTTCTATCGAGCGCCTCTCTTCTGGTCTGCGCATTAACAG
CGCTAAAGATGACGCTGCGGGCCAAGCGATTGCTAACCGCTTCACTTCTA
ACATCAAAGGTCTGACTCAGGCCGACGTAACGCCAACGACGGTATTTCT
CTGGCGCAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTTGCA
GCGTGTTTCGTGAAGTACCCTTCAGGCCACTACCGGTACTAAGTCTGATTC
TGACCTGTCTTCAATACAGGACGAAATCAAATCCCGTCTCGATGAAATTG
ACCGCGTATCCGGTCAGACTCAGTTCAACGGCGTTAATGTTCTTTCCAAAG
ATGGTTCAATGAAAATTCAGGTTGGTGCGAATGATGGTCAAACCTATCTCC
ATCGATCTGAAGAAAATTGATTCTTCAACTTTGGGGCTGAATGGCTTCTCA
GTTTCTAAAACTCTCTTAATGTCAGCAATGCTATCACATCTATCCCGCAA
GCCGCTAGCAATGAACCTGTTGATGTTAACTTCGGTGATACTGATGAGTCT
GCAGCAATCGCAGCCAAATTGGGGGTTTCCGATACGTCAAGCCTGTGCGT
GCACAACATCCTTGATAAAGATGGTAAGGCAACAGCTGATTATGTTGTTT
AGTCAGGTAAAGACTTCTATGCTGCTTCTGTTAATGCCGCTTCAGGTAAAG
TAACCTTAAACACCATTGATGTTACTTATGATGATTATGCGAACGGTGTTG
ACGATGCCAAGCAAACAGGTCAGCTGATCAAAGTTTCAGCAGATAAAGAC
GGCGCAGCTCAAGGTTTTGTACACTTCAAGGCAAAAACTATTCTGCTGGT
GATGCGGCAGACATTCTTAAGAATGGAGCAACAGCTCTTAAGTTAACTGA
TCTGAATTTAAGTGATGTTACTGATACTAATGGTAAGGTAACCACAACCTGC
GACTGAGCAATTTGAAGGTGCTTCAACTGAGGATCCGCTGGCGCTTCTGG
ATAAAGCTATTGCATCAGTCGACAAATTCCGGTCTTCTCTAGGTGCCGTGC
AGAACCGTCTCGATTCCGCTATCACCAACCTGAACAACACCACCACCAAC
CTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCGACTATGCGACCGAAGT
GTCCAACATGTGAAAGCGCAGATCATCCAGCAGGCAGGTAACTCCGTGC
TGTCTAAAGCGAACCAGGTACCGCAGCAAGTTCTGTCACTGTTACAAGGC
TAATGGCCTTAACCTGCCTGACCCCGCCACCGGCGGGGTTTTTCTGTCCG
CAATTTACCGATAACCCCAAATAACCCCTCATTTACCCCACTAATCGTCC
GATTAATAAACCTGCAGAAACGGATAATCATGCCGATAACTCATATAACG
CAGGGCTGTTTATCGTGAATTCCTCTATACCGCTGAAGGTGTAATGGATA
AACACTCGCTG

Figure 69

92/96

AACAGCCTCTCGCTGATCACTCAGAACAACATCAACAAAAACCAGTCTTC
AATGTCTACTGCCATTGAGCGTCTGTCTTCCGGTCTGCGTATCAACAGCGC
AAAAGATGACGCTGCTGGCCAGGCGATTGCCAACCGCTTCACCTCTAACA
TCAAAGGTCTGACTCAGGCAGCTCGTAACGCCAACGACGGTATCTCCGTT
GCACAGACCACTGAAGGCGCACTGTCTGAAATCAACAACAACCTGCAGCG
TATCCGTGAGCTGACTGTTCACTCTTCTACGGGTACTAACTCTGAATCCGA
TCTGAACTCAATCCAGGACGAAATTAATCCCGTCTGGACGAAATTGACC
GCGTATCCGGTCAGACCCAGTTCAACGGCGTGAACGTGCTGGCAAAAGAC
GGCTCCATGAAAATTCAGGTTGGCGCGAACGATGGTGAAACCATCACCAT
CGACCTGAAAAAAATTGACTCTTCTACTTTAAACCTGACTGGGTTTAA

Figure 70A

93/96

CTCAGTATGCTGTCACCGGCAGTACAGGTGCCGTAACCTTACGATCCAGAT
ACAGATCCTGCCGCGACTGGTGATATTGTTTCTGCTTATGTTGATGATGCA
GGTACATTGACAACTGATGCAAACAAAACCTGTAAAATATTATGCCCCACAC
TAATGGTAGCGTCACGAACGACAGTGGTTCAGCTATTTACGCAACTGAAG
CGGGCAAATTGACTACTGAAGCGTCTACAGCTGCTGAAACTACCGCTAAC
CCACTGAAAGCCCTGGACGATGCAATCAGCCAGATCGACAAATTCCGTTC
TTCTCTGGGTGCTGTACAGAACCGTCTGGATTCTGCGGTAACCAACCTGAA
CAACACCACCACCAACCTGTCTGAAGCGCAGTCCCGTATTCAGGACGCCG
ACTATGCGACCGAAGTGTCAAATATGTCTAAAGCGCAGATCATCCAGCAG
GCCGGTAACTCCGTGTTGGCTAAAGCTAACCAGGTTCCCTCAGCAGGTT

Figure 70B

94/96

AGCCTGTCGCTGTTGACCCAGAATAACCTGAACAAATCTCAGTCTTCTCTG
AGCTCCGCCATTGAGCGTCTCTCTTCTGGCCTGCGTATTAACAGTGCTAAA
GATGACGCAGCAGGTCAGGCGATTGCTAACCGTTTTACAGCAAATATTAA
AGGTCTGACTCAGGCTTCCCGTAACGCGAATGATGGTATTTCTGTTGCGCA
GACCACTGAAGGCGCGCTGAATGAAATTAACAACAACCTGCAGCGTGTAC
GTGAACTGACTGTTTCAGGCAACTAACGGTACTAACTCTGACAGCGATCTTT
CTTCTATCCAGGCTGAAATTACTCAACGTCTGGAAGAAATTGACCGTGTAT
CTGAG CAAACTCAGTTTAACGGCGTGAAAGTCCTTGCTGAAAAT

Figure 71

95/96

GCACGTTAGTTGTTAACGGTGCAACTTACGATGTTAGTGCAGATGGTAAA
ACGATAACGGGAGACTGCTTCTGGTAACAATAAAGTCATGTATCTGAGCAA
ATCAGAAGGTGGTAGCCCGATTCTGGTAAACGAAGATGCAGCAAAATCGT
TGCAATCTACCACCAACCCGCTCGAAACTATCGACAAAGCATTGGCTAAA
GTTGACAATCTGCGTTCTGACCTCGGTGCAGTACAAAACCGTTTCGACTCT
GCTATCACCAACCTTGGCAACACCGTAAACAACCTGTCTTCTGCCCCGTAGC
CGTATCGAAGATGCTGACTACGCGACCGAAGTGTCTAACATGTCTCGTGC
GCAGATCCTGCAACAAGCGGGTACCTCTGTTCTGGCGCAGGCTAACCAGA
CCACGCAGAACGTAC

Figure 72

96/96

FIGURE 73A

Sequence of the polylinker region of plasmid pTic99A:

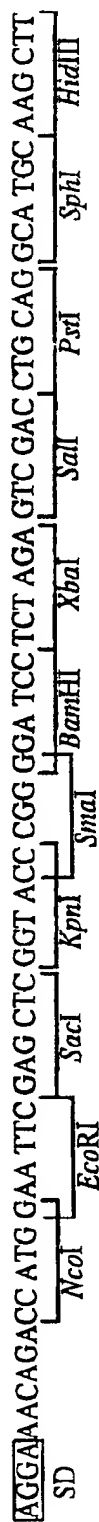
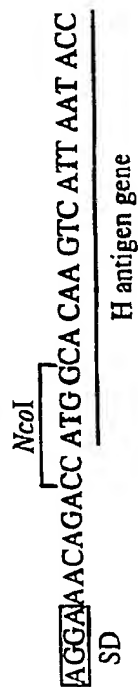


FIGURE 73B

Sequence in the junction region between vector and the 5' end of the H antigen gene:



SEQUENCE LISTING PART

<110> THE UNIVERSITY OF SYDNEY

<120> ANTIGENS AND THEIR DETECTION

<130> REEVES

<140>

<141>

<160> 68

<170> PatentIn Ver. 2.0

<210> 1

<211> 1773

<212> DNA

<213> Escherichia coli

<400> 1

```
atgCGacgta tagaacgaat accgggggta tcggcgtaag cgggggcaaag tttacgattt 60
atTTTTtggc ttaatgacac gaacagcaac gaggaagggg agtatttcga ccgctagaaa 120
aaaattctaa aggttgtagag tgaccagacg ataacagggg tgacggcgac gaagccgaag 180
ggtggaagcc caatacttaa accgtagact tgaaaacagg aaaatgaatc atggcacaag 240
tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag aaccagtctg 300
cgctgtcgac ttctatcgag cgctctctct ctggtctgcg cattaacagc gctaaagatg 360
acgctgcggg ccaagcgatt gctaaccgct tcacttctaa catcaaaggt ctgactcagg 420
ccgcacgtaa cgccaacgac ggtatttctc tggcgcgacac cactgaaggc gcactgtctg 480
aaatcaacaa caacttgtag cgtgttcgtg aactgaccgt tcaggccact accggtacta 540
actctgattc tgacctgtct tcaatacagg acgaaatcaa atcccgctct gatgaaattg 600
accgcgtatc cggtcagact cagttcaacg gcgttaatgt tctttccaaa gatggttcaa 660
tgaaaattca ggttggtgag aatgatggc aaactatctc catcgatctg aagaaaattg 720
attcttcaac tttggggctg aatggcttct cagtttctaa aaactctctt aatgtcagca 780
atgctatcac atctatcccg caagccgcta gcaatgaacc tgttgatggt aacttcgggtg 840
atactgatga gtctgcagca atcgagcca aattgggggt ttccgatacg tcaagcctgt 900
cgctgcacaa catccttgat aaagatggta aggcaacagc tgattatggt gttcagtcag 960
gtaaagactt ctatgctgct tctgttaatg ccgcttcagg taaagtaacc ttaaaccacca 1020
ttgatgttac ttatgatgat tatgcgaacg gtgttgacga tgccaagcaa acaggtcagc 1080
tgatcaaagt ttcagcagat aaagacggcg cagctcaagg ttttgtcaca cttcaaggca 1140
aaaactattc tgctgggtgat gcggcgagaca ttcttaagaa tggagcaaca gctcttaagt 1200
taactgatct gaatttaagt gatgttactg atactaatgg taaggtaacc acaactgcga 1260
ctgagcaatt tgaaggtgct tcaactgagg atccgctggc gcttctggat aaagctattg 1320
catcagtcga caaattccgg tcttctctag gtgcgctgca gaaccgtctc gattccgcta 1380
tcaccaacct gaacaacacc accaccaacc tgtctgaagc gcagtccgct attcaggacg 1440
ccgactatgc gaccgaagtg tccaacatgt cgaaagcgca gatcatccag caggcaggta 1500
actccgtgct gtctaaagcg aaccaggtag cgcagcaagt tctgtcactg ttacaaggct 1560
aatggcctta acctgcctga ccccgccacc ggcgggggtt tttctgtccg caatttaccg 1620
```

ataacccccca aataaccctt catttcaccc actaatcgtc cgattaaaaa ccctgcagaa 1680
 acggataatc atgccgataa ctcatataac gcagggtgtt ttatcgtaaa ttcactctat 1740
 accgctgaag gtgtaatgga taaacactcg ctg 1773

<210> 2

<211> 500

<212> DNA

<213> Escherichia coli

<400> 2

aacagcctct cgctgatcac tcagaacaac atcaacaaaa accagtcttc aatgtctact 60
 gccattgagc gtctgtcttc cggctctcgt atcaacagcg caaaagatga cgctgctggc 120
 caggcgattg ccaaccgctt cactctaac atcaaaggte tgactcaggc agctcgtaac 180
 gccaacgacg gtatctccgt tgcacagacc actgaaggcg cactgtctga aatcaacaac 240
 aacctgcagc gtatccgtga gctgactgtt cagtcttcta cgggtactaa ctctgaatcc 300
 gatctgaact caatccagga cgaaattaaa tcccgtctgg acgaaattga ccgcgtatcc 360
 ggtcagaccc agttcaacgg cgtgaacgtg ctggcaaaag acggctccat gaaaattcag 420
 gttggcgga acgatggtga aaccatcacc atcgacctga aaaaaattga ctcttctact 480
 ttaaacctga ctgggtttaa 500

<210> 3

<211> 500

<212> DNA

<213> Escherichia coli

<400> 3

ctcagtatgc tgtcaccggc agtacagggtg ccgtaactta cgatccagat acagatcctg 60
 ccgcgactgg tgatattggt tctgcttatg ttgatgatgc aggtacattg acaactgatg 120
 caaacaaaac tgtaaaatat tatgccaca ctaatggtag cgtcacgaac gacagtgggt 180
 cagctattta cgcaactgaa gcggggcaaat tgactactga agcgtctaca gctgctgaaa 240
 ctaccgctaa cccactgaaa gccctggacg atgcaatcag ccagatcgac aaattccgtt 300
 cttctctggg tgctgtacag aaccgtctgg attctgcggt aaccaacctg aacaacacca 360
 ccaccaacct gtctgaagcg cagtcccgta ttcaggacgc cgactatgag accgaagtgt 420
 caaatatgtc taaagcgag atcatccagc aggcgggtaa ctccgtgttg gctaaagcta 480
 accaggttcc tcagcaggtt 500

<210> 4

<211> 399

<212> DNA

<213> Escherichia coli

<400> 4

agcctgtcgc tgttgacca gaataacctg aacaaatctc agtcttctct gagctccgcc 60
 attgagcgtc tctcttctgg cctgcgtatt aacagtgtta aagatgacgc agcaggtcag 120
 gcgattgcta accgttttac agcaaatatt aaaggctctga ctcaggcttc ccgtaacgcg 180
 aatgatggta tttctgttgc gcagaccact gaaggcgcg cgaatgaaat taacaacaac 240
 ctgcagcgtg tacgtgaact gactgttcag gcaactaac gtactaactc tgacagcgat 300
 ctttcttcta tccaggctga aattactcaa cgtctggaag aaattgaccg tgtatctgag 360

caaactcagt ttaacggcgt gaaagtcctt gctgaaaat

399

<210> 5

<211> 417

<212> DNA

<213> Escherichia coli

<400> 5

gcacgttagt tgtaaacggt gcaacttacg atgttagtgc agatggtaaa acgataacgg 60
agactgcttc tggtaacaat aaagtcacgt atctgagcaa atcagaaggt ggtagcccca 120
ttctggtaaa cgaagatgca gcaaaatcgt tgcaatctac caccaaccgg ctcgaaaacta 180
tcgacaaaagc attggctaaa gttgacaatc tgcgttctga cctcgggtgca gtacaaaacc 240
gtttcgactc tgctatcacc aaccttggca acaccgtaaa caacctgtct tctgcccga 300
gccgtatcga agatgctgac tacgcgaccg aagtgtctaa catgtctcgt gcgcagatcc 360
tgcaacaagc gggtagctct gttctggcgc aggctaacca gaccacgcag aacgtac 417

<210> 6

<211> 950

<212> DNA

<213> Escherichia coli

<400> 6

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgtatt 60
aacagcgcta aagatgacgc cgcgggccag gcgattgcta accgctttac ttctaaccatc 120
aaaggctctga ctgagccgc acgtaacgcc aacgacggta tttctctggc gcagacggct 180
gaaggcgcgc tgcagagat taacaacaac ttgcagcgta ttcgtgaact gaccgttcag 240
gcctctaccg gcacgaactc tgattccgac ctgtcttcta ttcaggacga aatcaaatacc 300
cgtcttgatg aaattgaccg tgtatctggc cagaccagct tcaacgggtg gaacgtgctg 360
tcgaaaaacg attcgatgaa gattcagatt ggtgccaatg ataaccagac gatcagcatt 420
ggcttgcaac aaatcgacag taccactttg aatctgaaag gatttaccgt gtccggcatg 480
gcggatttca gcgcggcgaa actgacggct gctgatggta cagcaattgc tgctgcggat 540
gtcaaggatg ctgggggtaa acaagtcaat ttactgtctt aactgacac cgcgtctaac 600
agtactaaat atgcggtcgt tgattctgca accggtaaat acatggaagc cactgtagtc 660
attaccggta cggcggcggc ggtaactgtt ggtgcagcgg aagtggcggg agccgctaca 720
gccgatccgt taaaagcact ggatgccgca atcgctaaag tcgacaaatt ccgctcctcc 780
ctcggtgccg ttcaaaaccg tctggattct gcggtcacca acctgaacaa caccaccacc 840
aacctgtctg aagcgcagtc ccgtattcag gacgccgact atgcgaccga agtgtccaac 900
atgtcgaaag cgcagattat ccagcaggcg ggcaactccg tgctgtctaa 950

<210> 7

<211> 1212

<212> DNA

<213> Escherichia coli

<400> 7

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgtatt 60
aacagcgcta aagatgacgc cgcgggccag gcgattgcta accgcttcac ttctaaccatc 120
aaaggctctga ctgagccgc acgtaacgcc aacgacggta tctctctggc gcgagaccat 180

```

gaaggcgcgc tgtctgaaat caacaacaac ttgcagcgtg tgcgtgagtt gaccgttcag 240
gcgacgaccg ggactaactc tgattctgac ctgtcttcta ttcaggacga aatcaaatecc 300
cgtctggatg aaattgatcg cgtttccggt cagaccaggt tcaacggcgt gaatgtgctg 360
gcgaaagatg gttcgatgaa gattcaggtt ggcgcgaaat atgggcagac tattagcatt 420
gatttgcaga agattgactc ttctacatta ggactgaacg gtttctccgt ttcgggtcag 480
tcacttaacg ttagtgattc cattactcaa attaccggtg ccgccgggac aaaacctgtt 540
gggtgttgatt tcactgctgt tgcgaaagat ctgactactg cgacaggtaa aacagtcgat 600
gtttctagcc tgacgttaca caacactctg gatgcgaaag gggctgctac atcacagttc 660
gtcgttcaat ccggcaatga tttctactcc gcgtcgatta atcatacaga cggcaaagtc 720
acgttgaata aagccgatgt cgaatacaca gacaccgata atggactaac gactgcggct 780
actcagaaag atcaactgat taaagttgcc gctgactctg acggctcggc tgcgggatat 840
gtaacattcc aaggtaaaaa ctacgctaca acggtttcaa cggcacttga tgataatact 900
gcggcaaaaag caacagataa taaagttgtt gttgaattat caacagcaaa accgactgca 960
cagttctcag gggcttcttc tgctgatcca ctggcacttt tagacaaagc tattgcacag 1020
gttgatactt tccgctcctc cctcgggtgcg gtgcaaaacc gtctggattc cgcagtaacc 1080
aacctgaaca acaccaccac caacctgtct gaagcgcagt cccgtattca ggacgccgac 1140
tatgctacag aagtgtccaa catgtcgaaa gcgcagatca tccagcaggc aggtaaactcg 1200
gtgctgtcca aa
1212

```

<210> 8

<211> 1647

<212> DNA

<213> Escherichia coli

<400> 8

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
gcgctgtccg aaattaacaa caacttacag cgtattcgtg aactgacggt tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttggtgcg aatgacggcc agactatcac tattgatctg 480
aagaaaattg actctgatac gctggggctg aatgggttta atgtgaacgg caaaggggaa 540
acggctaata cggcagcaac cctgaaagat atgtctggat tcacagctgc ggcggcacca 600
gggggaactg ttggtgtaac tcaatatact gacaaatcgg ctgtagcaag tagcgtagat 660
attctaaatg ctggtgctgg cgcagatgga aataaagtta caactagcgc cgatgttggt 720
tttggtagac cagccgctgc tgtaacctat acctacaata aagacactaa ttcattattcc 780
gccgttctg atgatatttc cagcgctaac ctggctgctt tectcaatcc tcaggccgga 840
gatacgacta aagctacagt tacaattggt ggcaaagatc aagatgtaaa catcgataaa 900
tccggtaatt taactgctgc tgatgatggc gcagtacttt atatggatgc taccggtaac 960
ttaactaaaa ataattgctg tggtgatata caagctactt tggctaaact tgctactgct 1020
actggtgcta aagccgcgac catccaaact gataaaggaa cattcaccag tgacggtaca 1080
gcgtttgatg gtgcatcaat gtccattgat accaatacat ttgcaaatgc agtaaaaaat 1140
gacacttata ctgccactgt aggtgctaag acttatagcg taacaacagg ttctgctgct 1200
gcagacaccg ctatatgag caatgggggt ctcagtgata ctccgccaac ttactatgca 1260
caagctgatg gaagtatcac aactactgag gatcgggctg ccggtaaaact ggtctacaaa 1320
ggttccgatg gtaagttaac aacggatagc actagcaaaag cagaatcaac atcagatccg 1380

```

ctggcagctc ttgacgacgc tatcagccag atcgacaaat tccgctcctc cctgggtgcg 1440
gtgcaaaacc gtctggattc cgcagtgacc aacctgaaca acaccactac caacctgtct 1500
gaagcgcagt cccgtattca ggacgccgac tatgcgaccg aagtgtccaa catgtcgaaa 1560
gcgagatta tccagcaggc cggtaactcc gtgctggcaa aagctaacca ggttccgcag 1620
caggttctgt ctctgctgca ggggttaa 1647

<210> 9

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 9

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac caccgaaggc 240
gcgctgtctg aaatcaacaa caacttacag cgtatccgtg agctgacggt tcaggcttct 300
accggaacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctt 360
gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttggtgcg aatgacggtg aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaattg taaagggtact 540
attaccaaca agctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaacacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcggcg tagattatac ttacaacgct 720
aaatctggtg attttactac caccaaactc actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaatctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cgggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct taaagccgcg 1020
agcgaaggta gtgacggtgc ttctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtggtatta ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaattg gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgagatt 1680
atccagcagg ccggtaaactc cgtgctggca aaagccaacc aggtaccgca gcaggttctg 1740
tctctgctgc aggggttaa 1758

<210> 10

<211> 1383

<212> DNA

<213> Escherichia coli

<400> 10

aacaaatctc agtcttctct tagctctgct attgagcgtc tgtcttctgg tctgcgatt 60
aacagcgcaa aagacgatgc agcaggctcag gcgattgcta accgttttac ggcaaattatt 120
aaagggtctga cccaggcttc ccgtaacgca aatgatggta tttctgttgc gcagaccact 180
gaagggtgcgc tgaatgaaat taacaacaac ctgcagcgta ttcgtgaact ttctgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggtt ggtgctaatt atggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggctggacg gttttaatat cgatggcgcg 480
cagaaagcaa caggcagtgta cctgatttct aaatttaaag cgacaggtag tgataattat 540
gatgttggcg gtaaaactta taccgtgaat gtggagagcg gcgcggttaa gaatgatgct 600
aataaagatg tttttgtaag cgcagctgat ggatcgctga cgaccagtag tgatactaaa 660
gtatccggtg aaagtattga tgcaacagaa ctacgcaaac ttgcaataaa attagctgac 720
aaaggctcca ttgaatacaa gggcattaca ttactaaca aactggcgcg agagcttgat 780
gctaattggtg aagggtgtttt gaccgcaaat attgatggtc aagatgttca atttactatt 840
gacagtaatg caccacggtg tgccggcgca acaataacta cagacacagc tgtttacaaa 900
aacagtgcgg gccagttcac cactacaaaa gtggaaaata aagccgcaac actctctgat 960
ctggatctta atgcagccaa gaaaacaggt agcactttag ttgtaaattg cgccacctac 1020
aatgtcagcg cagatggtaa aacggtaact gatactactc ctggtgcccc taaagtgatg 1080
tatctgagca aatcagaagg tggtagcccc attctggtaa acgaagatgc agcaaaatcg 1140
ttgcaatcta ccaccaaccc gctcgaaact atcgacaagg cattggctaa agttgacaat 1200
ctgcgttctg acctcggtgc agtacaaaac cgtttcgact ctgccatcac caaccttggc 1260
aacaccgtaa acaacctgtc ttctgcccgt agccgtatcg aagatgctga ctacgcgacc 1320
gaagtgtcta acatgtctcg tgcgagatc ctgcaacaag cgggtacctc tgttctggcg 1380
cag 1383

<210> 11

<211> 2013

<212> DNA

<213> Escherichia coli

<400> 11

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttccg ttgcacagac cactgaaggc 240
gcgctgtccg aaattaacaa caacttacag cgtattcgtg aactgacggt tcaggcttct 300
accgggacta actccgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc cgccagacc cagttcaacg gcgtgaacgt gctgtccaaa 420
gatggctcga tgaaaattca ggtcggcgcg aacgatggcg aaacgattac tattgatctg 480
aagaaaattg actctgatac gctgaatctg gctgggttta acgttaacgg taaaggttct 540
gtagcgaata cagctgcgac aagcgacgat taaaactgg ctgggtttcac taagggcacc 600
acagatacca atggcgtgac cgcgtatata aacacaatta gtaatgacaa agccaaagct 660
tccgatctgt tagctaatat caccgatgga tcagtgatca ctgggggagg ggcaaacgct 720
tttggcgtgg ctgcaaagaa tggttacacc tatgatgcag caagtaaate ttatagtttt 780
gctgcagatg gtgccgattc agcgaagacg ttaagcatca ttaatccaaa caccggtgat 840
tcgtcgcagg cgacagtgac tattgggtgg aaagagcaga aagttaatat ttcccaggat 900
ggaaaaatta ctgcggcaga tgataatgcg acgctgtatt tagataaaca gggaaacttg 960

acaaaaacga atgcaggtaa cgataccgca gcgacttggg atggtttaat ttccaacagc 1020
gattctaccg gtgcgggtcc agttgggggtt gcaactacaa ttacaattac ttctggtaca 1080
gcttcgggaa tgtctgttca gtccgcagga gcaggaattc agacctcaac aaattctcag 1140
attcttgcag gtggtgcatt tgcggctaag gtaagtattg agggaggcgc tgctacagac 1200
atthtggtag caagtaatgg aaacataaca gcggctgatg gtagtgcact ttatcttgat 1260
gcgactactg gtggattcac tacaacggct ggaggaaata cagctgcttc gttagataat 1320
ttaattgcta acagtaagga tgctacctta accgtaactt caggtaccgg ccagaacact 1380
gtttatagca caacaggaag tggcgctcag ttcaccagtt tagcaaaagt agacacagtc 1440
aatgtcacca acgcacatgt cagtgccgaa ggtatggcaa atctgacaaa aagcaatttt 1500
accattgata tgggcgggtac aggtacagta acttacacag tttccaatgg ggatgtgaaa 1560
gctgctgcaa atgctgatgt ttatgtcgaa gatgggtgcac tttcagccaa tgctacaaaa 1620
gatgtaacct actttgaaca aaaaaatggg gctattacca acagcaccgg tggtagcatc 1680
tatgaaacag ctgatggtaa gttacaaca gaagctacta ctgcatccag ttccaccgcc 1740
gatccctga aagctctgga cgaagccatc agctccatcg acaaattccg ctctctcctc 1800
gggtgcgggtgc aaaaccgtct ggattccgcg gtcaccaacc tgaacaacac cactaccaac 1860
ctgtccgaag cgcagtcccg tattcaggac gccgactatg cgaccgaagt gtccaacatg 1920
tcgaaagcgc agatcatcca gcaggccggt aactccgtgc tggcaaaagc taaccaggta 1980
ccgcagcagg ttctgtctct gctgcagggt taa 2013

<210> 12

<211> 1263

<212> DNA

<213> Escherichia coli

<400> 12

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
gcgctgtccg aaattaacaa caacttacag cgtgtgcgtg agctgactgt tcaggcgacc 300
accggtacta actctgagtc tgacctgtct tctatccagg acgaaatcaa atctcgctg 360
gaagagattg atcgtgtttc aagtcagact caatttaacg gcgtgaatgt tttggctaaa 420
gatgggaaaa tgaacattca ggttggggca aatgatggac agactatcac tattgatctg 480
aaaaagatcg attcatctac actaaacctc tccagttttg atgtacaaa cttgggcacc 540
agtgttaaaag atggggccac catcaataag caagtggcag taggtgctgg cgactttaaa 600
gataaagctt caggatcgtt aggtacccta aaattagttg agaaagacgg taagtactat 660
gtaaatgaca ctaaaagtag taagtactac gatgccgaag tagatactag taagggtaaa 720
attaacttca actctacaaa tgaaagtgga actactccta ctgcagcgac ggaagtaact 780
actgttggcc gcgatgtaaa attggatgct tctgcactta aagccaacca atcgcttgct 840
gtgtataaag ataaaagcgg caatgatgct tatatcattc agaccaaaga tgtaacaact 900
aatcaatcaa ctttcaatgc cgctaataatc agtgatgctg gtgttttattc tattggtgca 960
tctacaaccg cgccaagcaa tttaacagct aacccgctta aggtctttga tgatgcaatt 1020
gcatctgttg ataaattccg ctcttctctc ggtgccgttc agaaccgtct ggattctgcc 1080
attgccaaac tgaacaacac cactaccaac ctgtctgaag cgcagtcccg tattcaggac 1140
gctgactatg cgaccgaagt gtccaacatg tcgaaagcgc agattatcca gcaggccggt 1200
aactccgtgc tggcaaaagc caaccaggta ccgcagcagg ttctgtctct gctgcagggt 1260
taa 1263

<210> 13

<211> 1368

<212> DNA

<213> Escherichia coli

<400> 13

```

aacaaatctc agtcttctct gagctccgcc attgaacgtc tctcttcttg cctgcgtatt 60
aacagtgcta aagatgacgc agcaggctcag gcgattgcta accgttttac agcaaattatt 120
aaagggtctga ctcaggcttc ccgtaacgcg aatgatggta tttctgttgc gcagaccact 180
gaagggtgcgc tgaatgaaat taacaacaac ctgcagcgtg tacgtgaact gactgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggtt ggtgctaata atgggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggcctggacg gttttaatat cgatggcgcg 480
cagaaagcaa ctggcagtg cctgatttct aaatttaaag cgacaggtag tgataactat 540
gatgttggcg gtgatgctta tactgttaac gtagatagcg gagctgttaa agatactaca 600
gggaatgata tttttgttag tgcagcagat ggttcactga caactaaatc tgacacaaac 660
atagctggta cagggattga tgctacagca ctgcagcag cggctaagaa taaagcacag 720
aatgataaat tcacgtttaa tggagttgaa ttcacaacaa caactgcagc ggatggcaat 780
gggaatggtg tatattctgc agaaattgat ggtaagtcag tgacatttac tgtgacagat 840
gctgacaaaa aagcttcttt gattacgagt gagacagttt acaaaaatag cgctggcctt 900
tatacgacaa ccaaagttga taacaaggct gccacacttt ccgatcttga tctcaatgca 960
gctaagaaaa caggaagcac gttagttgtt aacggtgcaa cttacgatgt tagtgacagat 1020
ggtaaaacga taacggagac tgcttctggt aacaataaag tcatgtatct gagcaaataca 1080
gaagggtgga gcccgattct ggtaaacgaa gatgcagcaa aatcggttga atctaccacc 1140
aaccgcgtcg aaactatcga caaagcattg gctaaagttg acaatctgcg ttctgacctc 1200
ggtgcagtac aaaaccgttt cgactctgct atcaccaacc ttggcaaacac cgtaaacaac 1260
ctgtcttctg cccgtagccg tatcgaagat gctgactacg cgaccgaagt gtctaacatg 1320
tctcgtgcgc agatcctgca acaagcgggt acctctgttc tggcgagc 1368

```

<210> 14

<211> 1788

<212> DNA

<213> Escherichia coli

<400> 14

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccgtt tcacctctaa cattaaaggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtatctccg ttgcgcagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggg tcaggcttct 300
accgggacta actccgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc tggccagacc cagttcaacg gcgtgaacgt actggcgaaa 420
gacggttcaa tgaaaattca ggttggtgag aatgacggcc agactatcac gattgatctg 480
aagaaaattg actcagatac gctggggctg aatgggttta acgtgaatgg ttccggtacg 540
atagccaata aagcggcgac cattagcgac ctgacagcag cgaaaatgga tgctgcaact 600
aatactataa ctacaacaaa taatgcgtg actgcatcaa aggcgcttga tcaactgaaa 660
gatggtgaca ctgttactat caaagcagat gctgctcaaa ctgccacggt ttatacatat 720

```

```

aatgcatcag ctggtaactt ctcattcagt aatgtatcga ataatacttc agcaaaagca 780
ggtgatgtag cagctagcct tctcccgccg gctgggcaaa ctgctagtgg tgtttataaa 840
gcagcaagcg gtgaagtga ctttgatggt gatgcgaatg gtaaaatcac aatcggagga 900
cagaaagcat atttaactag tgatggtaac ttaactacaa acgatgctgg tgggtgcgact 960
gcggtacgc ttgatggttt attcaagaaa gctgggtgatg gtcaatcaat cgggtttaag 1020
aagactgcat cagtcacgat ggggggaaca acttataact ttaaaacggg tgctgatgct 1080
gatgctgcaa ctgctaacgc aggggtatcg ttcactgata cagctagcaa agaaaccgtt 1140
ttaaataaag tggctacagc taaacaaggc aaagcagttg cagctgacgg tgatacatcc 1200
gcaacaatta cctataaatc tggcgttcag acgtatcagg ctgtatttgc cgcaggtgac 1260
ggtactgcta gcgcaaaata tgccgataaa gctgacgttt ctaatgcaac agcaacatac 1320
actgatgctg atggtgaaat gactacaatt ggttcataca ccacgaagta ttcaatcgat 1380
gctaacaacg gcaaggtaac tgttgattct ggaactggta cgggtaaata tgcgccgaaa 1440
gtaggggctg aagtatatgt tagtgctaat ggtactttaa caacagatgc aactagcgaa 1500
ggcacagtaa caaaagatcc actgaaagct ctggatgaag ctatcagctc catcgacaaa 1560
ttccgttctt ccctgggtgc tatccagaac cgtctggatt ccgcagtcac caacctgaac 1620
aacaccacta ccaacctgtc cgaagcgcag tcccgtattc aggacgccga ctatgcgacc 1680
gaagtgtcca acatgtcgaa agcgcagatc attcagcagg ccggtaaactc cgtgctggca 1740
aaagccaacc aggtaccgca gcaggttctg tctctgctgc aggggttaa 1788

```

<210> 15

<211> 1653

<212> DNA

<213> Escherichia coli

<400> 15

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttccg ttgcccagac cactgaagg 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg agctgacggg tcaggcttct 300
accgggacta actccgattc tgacctggac tccatccagg acgaaatcaa gtctcgtctg 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacgggtcga tgaaaattca gggttggtgcg aatgacggcc agactatcac gattgatctg 480
aagaaaattg actcagatac gctggggctg agtgggttta atgtgaatgg tggcggggct 540
gttgctaaca ctgctgcatac taaagctgac ttggtagctg ctaatgcaac tgtggtaggc 600
aacaaatata ctgtgagtg cgggttacgat gctgctaaag cgtctgattt gctggctgga 660
gttagtgatg gtgatactgt tcaggcaacc attaataacg gcttcggaac ggcggctagt 720
gcaacgaatt acaagtatga cagtgcaggt aagtcttact cttttgatac cacaacggct 780
tcagctgccg atgttcagaa atatttgacc ccgggcgttg gtgataccgc taagggcact 840
attactatcg atggttctgc acaggatgtt cagatcagca gtgatggtaa aattacgtca 900
agcaatggag ataaacttta cattgataca actgggcgct taacgaaaaa cggctttagt 960
gcttctttga ctgaggctag tctgtccaca cttgcagcca ataataccaa agcgacaacc 1020
attgacattg gcggtacctc tatctccttt accggtaata gtactacgcc gaacactatt 1080
acttattcag taacagggtgc aaaagttgat caggcagctt tcgataaagc tgtatcaacc 1140
tctggaaacg atgttgattt cactaccgca gggttatagc tcgacggcgc aactggcgct 1200
gtaacaaaag gtgttgctcc gggttatatt gataacaacg gggcggtgac cacatctgat 1260
actgtagatt tttatctaca ggatgatggt tcagtgacta acggcagcgg taaggcagtt 1320
tataaagatg ctgacggtaa attgacgaca gatgctgaaa ctaaagctgc aaccaccgcc 1380

```

gatccctga aagctctgga cgaagccatc agctccatcg acaaattccg ctctccctc 1440
ggtgcggtgc agaaccgtct ggattccgcg gtcaccaacc tgaacaacac cactaccaac 1500
ctgtctgaag cgcagtcctg tattcaggac gctgactatg cgaccgaagt atccaacatg 1560
tcgaaagcgc agatcatcca gcaggccggt aactccgtgc tggcaaaagc taaccaggta 1620
ccacagcagg ttctgtctct gctgcagggt taa 1653

<210> 16

<211> 1689

<212> DNA

<213> Escherichia coli

<400> 16

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtgtgctg aactgaccgt tcaggcaacc 300
accggtacca actcccagtc tgacctggac tctatccagg acgaaattaa atcccgtctg 360
gacgaaattg atcgcgatc cggtcagacc cagttcaacg gcgtgaacgt gctggcaaaa 420
gacggttcca tgaaaattca ggttggcgcg aacgatggcc agaccatcac tatcgacctg 480
aagaagattg actcttctac cttgaacctg acaggtttta acgttaacgg ttctggttct 540
gtggcggaata ctgcagcaac taaagctgat ttaaccgctg ctcaactctc tgcaccgggt 600
gcagcagacg caaatggtac agttacttat actgtcagtg ctgggtataa agaatccact 660
gctgcagatg ttattgctag catcaaagac ggagtgctc cgacttctgc aattactgca 720
accattaata atggcttcgg tgattccagt gcgctgactt ccaatgacta tacttatgac 780
ccagcaaaaag gcgacttcac ttacgacgta gcttcaagcg ccaataatac tgctgccag 840
gttcagtcct tcctgacgcc gaaagcaggt gataccgcaa atctgaaagt aaccgttggt 900
acgacatcgg ttgatgtcgt tctggccagt gatggtaaga ttacagcaaa agatggttct 960
gcattatata tcgacagtac aggtaacctg actcagaaca gtgctggctt gacctctgct 1020
aaactggcta ctctgactgg ccttcagggc tctggtgttg cttcaaccat cactactgaa 1080
gatggcacta atattgatat tgctgctaac ggtaatatgt gtctgaccgg tgttcgtatc 1140
agtgtgatt ctctgcagtc agcgactaaa tctacgggct ttactgttgg tactggcgct 1200
acaggtctga ccgtaggtac tgatggtaaa gtgactatcg gcgggactac tgctcagtc 1260
tacaccagca aagatggttc cctgactact gataacacca cttaaactgta tctgcagaaa 1320
gatggctctg taaccaacgg ttcaggtaaa gcggtctatg tagaagcgga tgggtgatttc 1380
actaccgacg ctgcaaccaa agccgcaacc accaccgatc cgctgaaagc cctggatgag 1440
gcaatcagcc agatcgataa gttccgttca tccctgggtg ctatccagaa ccgtctggat 1500
tccgcggtca ccaacctgaa caacaccact accaacctgt ctgaagcgca gtcccgtatt 1560
caggacgccg actatgcgac cgaagtgtcc aacatgtcga aagcgcagat cattcagcag 1620
gccggttaact ccgtgctggc aaaagccaac caggtaccgc aacaggttct gtctctgctg 1680
cagggctaa 1689

<210> 17

<211> 915

<212> DNA

<213> Escherichia coli

<400> 17

gcgctgtcga cttctatcga gcgcctctct tctgggtctgc gtattaacag cgctaaagat 60
gacgctgcgg gccaggcgat tgctaaccgc ttcacttcta acatcaaagg tctgactcag 120
gccgcacgta acgccaacga cgggtatttct ctggcgcgaga cggctgaagg cgcgctgtca 180
gagattaaca acaacttgca gcgtattcgt gaactgaccg ttcaggcctc taccggcacg 240
aactctgatt ccgacctgtc ttctattcag gacgaaatca aatcccgtct tgatgaaatt 300
gaccgtgtat ctggtcagac ccagttcaac ggtgtgaacg tgctgtcgaa aaacgattcg 360
atgaagattc agattgggtgc caatgataac cagacgatca gcattggctt gcaacaaatc 420
gacagtacca ctttgaatct gaaaggattt accgtgtccg gcatggcgga tttcagcgcg 480
gcgaaactga cggctgctga tggtagcaga attgctgctg cggatgtcaa ggatgctggg 540
ggtaaacaag tcaatttact gtcttacct gacaccgct ctaacagtac taaatatgcg 600
gtcgttgatt ctgcaaccgg taaatacatg gcagccactg tagtcattac cagtacggcg 660
gcggcggtaa ctggtgggtgc aacggaagtg gcgggagccg ctacagccga accgttaaaa 720
gcactggatg ccgcaatcgc taaagtcgac aaattccgct cctccctcgg tgccgttcaa 780
aaccgtctgg attctgcggc caccaacctg aacaacacca ccaccaacct gtctgaagcg 840
cagtcgccgta ttcaggacgc cgactatgcg accgaagtgt ccaacatgtc gaaagcgcg 900
attatccagc aggcg 915

<210> 18

<211> 1665

<212> DNA

<213> Escherichia coli

<400> 18

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa tattaaaggc 180
ctgactcagg ctgcacgtaa cgccaatgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aatcaacaa caacttacag cgtattcgtg aactgacggc tcaggccact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctg 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt gctgtccaaa 420
gatggttcaa tgaaaattca ggtcggcgca aatgatgggt aaaccatcac gattgatctg 480
aagaaaattg actctgatac gctgaatctg gctggtttta acgtgaatgg cgaagggtgaa 540
acagccaata ctgctgcaac acttaaagat atggttggtt taaaactcga taatacgggg 600
gtcactacag ctggagttaa tagatatatt gctgacaaag ccgtcgcaag tagcacggat 660
attttgaatg cggtagctgg tgttgatggc agtaaagttt ccacggaggc agatgttggt 720
tttggtgcag ctgcccctgg tacgccagtg gaataactt atcataaaga tactaacaca 780
tatacggctt ctgcttcagt tgatgcgact caactggcgg cattcctgaa tctgaagcg 840
ggtggtacca ctgctgcaac agtaagtatt ggcaacggta caacagctca agagcaaaaa 900
gtcattattg ctaaagatgg ttctttaact gctgctgatg acggtgccgc tctctatctt 960
gatgatactg gtaacttaag taaaactaac gcaggcactg atactcaagc taaactgtct 1020
gacttaatgg caaacaatgc taatgccaaa acagtcatta caacagataa aggtacattt 1080
actgetaata cgacaaagtt tgatggggta gatatttctg ttgatgcttc aacgtttgct 1140
aacgccgtta aaaatgagac ttacactgca actggttggtg taactttacc tgcgacatat 1200
acagtcaata atggcactgc tgcacagcg tatttagtcg atggaaaagt gagcaaaact 1260
cctgccgagt attttgcctc agctgatggc actattacta gtgggtgaaaa tgcggctacc 1320
agtaaagcta tctatgtaag tgccaatggc aacttaacga ctaatacaac tagtgaatct 1380
gaagctacta ccaaccgct ggcagcattg gatgacgcta tcgctctat cgacaaattc 1440
cgttcttccc tgggtgctat ccagaaccgt ctggattccg cagtcaccaa cctgaacaac 1500

accactacca acctgtctga agcgcagtc cgtattcagg acgccgacta tgcgaccgaa 1560
gtgtccaaca tgtcgaaagc gcagatcatt cagcaggccg gtaactccgt gctggcaaaa 1620
gccaaccagg taccgcagca gggtctgtct ctgctgcagg gttaa 1665

<210> 19

<211> 1842

<212> DNA

<213> Escherichia coli

<400> 19

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac cactgaaggc 240
gcgctgtccg aaattaacaa caacttacag cgtattcgtg aactgacggt tcaggcgacg 300
accggaacta actccacctc tgacctggac tccatccagg acgaaatcaa atcccgtctt 360
gacgaaattg accgcgtatc tggtcagacc cagttcaacg gcgtgaacgt gctgtctaaa 420
gatggctcga tgaaaattca ggtcggcgcg aacgatggcg aaacgattac tattgatctg 480
aagaaaattg actctgatac gctgaatctg gctggtttta acgttaacg taaaggttct 540
gtagcgaata ccgctgcgac tacagataat ctgacattgg ctgggttttac agcgggtact 600
aaagctgctg atggcacctg aacttatagc aaaaatgtcc agtttgccgc cgcgactgca 660
agcaatgtac tggctgctgc taaagatggc gacgaaatta cgttcgctgg taataacggc 720
acaggtatag ctgcaactgg ggggacttat acttatcata aggactctaa ctcatcacg 780
tttagcgcaa cggctgcac taaagattct ctgttgagca cactggcacc aaacgctggc 840
gatacattta ccgctaaagt gactattggg tctaaatcgc aagaagttaa cgtagcaaaa 900
gatggtacga ttacatccag cgatggtaag gcgctgtatt tagatgagaa gggcaacctg 960
acccaaacag gtagtggcac aaccaaagct gcaacctggg ataacctgat ggccaatata 1020
gatactacag gcaaagatgc ctatggtaac tctgcggcag cagctgttgg gacagtaatc 1080
gaagcaaaaag gaatgaccat cacttctgct ggtggtaatg ctcaggtgtt aaaagacgcg 1140
gcttataatg ccgcatatgc gacctcaatt actactggta ctccgggtga tgcgggagcc 1200
gcgggagccg ctgcaactgc gggtaatgcc gcggtgggag cgctgggcgc aacggcagtt 1260
gataatacca cggcagatgt tgccgatatc tctatctcag ctccgcaa atggcgagcatc 1320
cttcaggata aagatttcac cttaagtgat ggtagtgata cttacaacgt gaccagcaat 1380
gctgtcacta tcaatggcaa agcagcaaac attgatgaca gcggcgcaat cacagaccaa 1440
accagtaaag ttgtcaatta tttcgctcat actaacggtg gcgtgactaa cgatacaggc 1500
tccactattt atgcgacaga agatggtagc ctgaccaccg atgcagcaac caaagccgaa 1560
accaccgcg atcccctgaa agctctggac gaagccatca gctccatcga caaattccgc 1620
tcctccctcg gtgcgggtgca aaaccgtctg gattccgcgg tcaccaacct gaacaacacc 1680
accaccaacc tgtctgaagc gcagtcccgt attcaggacg ccgactatgc gaccgaagtg 1740
tccaacatgt cgaaagcgca gattatccag caggccggta actccgtgct ggcaaaaagt 1800
aaccaggtac cacagcaggt tctgtctctg ctgcagggtt aa 1842

<210> 20

<211> 1731

<212> DNA

<213> Escherichia coli

<400> 20

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaagggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
gcgctgtccg aaattaacaa caacttacag cgtgtgcgtg agctgactgt tcaggcgacc 300
accggtacca actcccagtc tgatctggac tctatccagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt gctggcaaaa 420
gacggttcca tgaaaattca ggttggcgcg aatgatggcc agaccatcac tatcgacctg 480
aagaagattg actcttctac gttgaaactg actggtttta acgtgaatgg ttctggttct 540
gtggcgaata ctgccccgac taaagcggat ttggctgctg ctgcaattgg taccctggg 600
gcagcagatt ctacaggtgc cattgcttac acagtaagtg ctgggctgac taaaactaca 660
gccgcagatg tactgtctag cctcgctgat ggtacgacta ttacagccac aggcgtgaaa 720
aatggctttg ctgcaggagc cacttccaat gcctataaac ttaacaaaga taataataca 780
tttacttatg acacgactgc tacgacagct gagctgcagt cttacctgac tccgaaagcg 840
ggcgacactg caacattcag tgttgaaatt ggtggtacta cacaagacgt cgtgctgtcc 900
agtgatggca aactcactgc taaggatggc tctaagcttt acattgatac aactggtaat 960
ttaactcaga atggtggtaa taacggtgtt ggaacactcg cggaagcgac tctgagtgg 1020
ttagctctga acaaaaatgg tttaacggct gttaaatcca caattactac agctgataac 1080
acttcgattg tactgaatgg ttcaagcgat ggtactggta atgctggtac tgaaggtagc 1140
attgctgtta caggcgctgt aattagttca gctgctctgc aatctgcaag caaaacgact 1200
ggtttcactg ttggtacagt agacacagct ggttatatct ctgtaggtac tgatgggagt 1260
gttcaggcat atgatgctgc gacttctggc aacaaagctt cttacaccaa cactgacggt 1320
acactgacta ctgataaacac cactaaactg tatctgcaga aagatggctc tgtaaccaac 1380
ggttcaggta aagcggctca tgtagaagcg gatggtgatt tcaactaccg cgctgcaacc 1440
aaagccgcaa ccaccaccga tccgctggcc gctctggatg acgcaatcag ccagatcgac 1500
aagtccggtt catccttggg tgctatccag aaccgtctgg attctgcagt caccaacctg 1560
aacaacacca ccaccaacct gtctgaagcg cagtcocgta ttcaggacgc cgactatgag 1620
accgaagtgt ccaatatgtc gaaagcgcag atcatccagc aggcgggtaa ctccgtgctg 1680
gcaaaagcca accagggtacc gcagcaggtt ctgtctctgc tgcagggtta a 1731

```

<210> 21

<211> 1380

<212> DNA

<213> Escherichia coli

<400> 21

```

aacaaatctc agtcttctct gagctccgcc attgaaagtc tctcttcttg cctgcgtatt 60
aacagtgcta aagatgacgc agcaggctcag gcgattgcta accgttttac agcaaattatt 120
aaaggctctga ctcaggcttc ccgtaacgcg aatgatggta tttctgttgc gcagaccact 180
gaagggtgagc tgaatgaaat taacaacaac ctgcagcgta ttcgtgaact ttctgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggt ggtgctaata atggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggcctggacg gttttaatat cgatggcgcg 480
cagaaagcaa ccggcagtga cctgatttct aaatttaaag cgacaggtag tgataattat 540
caaattaacg gtactgataa ctatactgtt aatgtagata gtggcgtagt acaggataaa 600
gatggcaaac aagtttatgt gagtactgag gatggttcac ttacgaccag cagtgatact 660
caattcaaga ttgatgcaac taagcttgca gtggctgcta aagatttagc tcaagggaat 720

```

```

aagattgtct acgaaggtat cgaatttaca aataccggca ctgtcgctat agatgccaaa 780
ggtaatggta aattaaccgc caatgttgat ggtaaggctg ttgaattcac tatttcgggg 840
agtactgata catcaggtac tagtgcaacc gttgccctta cgacagccct atacaaaaat 900
agtgcagggc aattgactgc aacaaaagtt gaaaataaag cagcgacact atctgatctt 960
gatctgaacg ctgccaaagaa aacaggaagc acgttagttg ttaacgggtgc aacttacgat 1020
gttagtgacg atggtaaaac gataacggag actgcttctg gtaacaataa agtcatgtat 1080
ctgagcaaat cagaagggtg tagcccgatt ctggtaaacy aagatgcagc aaaatcggtg 1140
caatctacca ccaacccgct cgaaactatc gacaaagcat tggctaaagt tgacaatctg 1200
cgttctgacc tcgggtgcagt acaaaaccgt ttcgactctg ccatcaccaa ccttggaac 1260
accgtaaaca acctgtcttc tgcccgtagc cgtatcgaag atgctgacta cgcgaccgaa 1320
gtgtctaaca tgtctcgtgc gcagatcctg caacaagcgg gtacctctgt tctggcacag 1380

```

<210> 22

<211> 1767

<212> DNA

<213> Escherichia coli

<400> 22

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccggt ttacttctaa cattaaggc 180
ctgactcagg cggcacgtaa cgccaacgac ggtatctctc tggcgagac caccgaagg 240
gcgctgtctg aaatcaacaa caacttacag cgtgtacgtg aactgaccgt tcaggcaacc 300
accggtacta actccgactc cgacctggct tctattcagg acgaaatcaa atcccgctctg 360
gatgaaattg accgcgtatc tggtcagact cagttcaacg gcgtgaacgt gctggcaaaa 420
gacggttcca tgaaaattca ggtaggtgct aacgacggcc agactatcac tattgacctg 480
aaaaaaatcg actctgatac tctgggcctg aatgggttta acgtgaatgg ttctgggacg 540
attaccaaca aagcagcaac tgtcagtgat gttactcgcg caggcggtac attggtgaat 600
ggtgcctatg atataaaaac cactaacaca gcgctgacta caactgatgc ctctcgcaaa 660
ttgaatgatg gtgatgttgt tactatcaat aatggtaagg atactgccta taaatataat 720
gctgctacag gtgggtttac gacggatgtc tccatctccg gggatcctac cgctgctgac 780
gctactgcta ataaaactgc ccgtgatgca cttgcggcgt ctttacatgc tgagccgggt 840
aaaactgtta atggttcttg gactacgaat gatggtacgg taaaatttga taccgatgcc 900
gatggaaga ttctatttgg tgggtgtgct gcttatgtag atgcagcagg caacctgacc 960
actaacgcag caggatgatg gactcaagca acaactaccg atttggttac tgctgctgca 1020
tctgctactg gtaagggtgg atccctgacc tttggtgaca cgacgtataa aattggctcag 1080
ggtacggctg gggttgatcc tgatgacgct tcagatgatg tactgggcac catttcttac 1140
tctaaatcag taagcaagga tgttgttctt gctgatacta aagcaactgg taacacgaca 1200
acagttgatt tcaactccgg tatcatgact tcaaagggtta gtttcgatgc aggtacatca 1260
actgatacat tcaagatgac agatggtgct atcaccaaaa cttaaagaata caccacttct 1320
tatgctgtaa ataaagatac tgggtgaagtt accgttgctg attatgctgc ggtagatagc 1380
gccgataagg ctgttgatga tactaaatata aaaccgacta tcggcgcgac agttaacctg 1440
aattctgcag gtaaatgtac cactgatacc accagtgcag gcacagcaac caaagatcct 1500
ctggctgccc tggacgctgc tatcagctcc atcgacaaat tccgttcac cctgggtgct 1560
atccagaacc gtctggattc cgcagtcacc aacctgaaca acaccactac caacctgtcc 1620
gaagcgcagt cccgtattca ggacgccgac tatgcgaccg aagtgtccaa catgtcga 1680
gcgcagatta tccagcaggc cggtaactcc gtgctggcaa aagccaacca ggtaccgcag 1740
caggttctgt ctctgctaca ggggttaa 1767

```

- 15 -

<210> 23

<211> 1383

<212> DNA

<213> Escherichia coli

<400> 23

```
aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tttcttctgg tctgcgtatt 60
aacagcgcta aagatgacgc tgcgggccag gcgattgcta accgcttcac ttctaacatc 120
aaaggctcga ctcaggccgc acgtaacgcc aacgacggta tttctctggc gcagaccact 180
gaaggcgcg cgtctgagat taacaacaac ttgcagcgtg tgcgtgagtt gactgtacag 240
gcgacgaccg ggactaactc tgattctgac ctgtcttcta tccaggatga aatcaaatcc 300
cgtttaagcg aaattgaccg tgtatctggg cagactcagt ttaacggcgt gaacgtactg 360
gctaagaatg acaccctgtc tattcaggta ggtgcaaagt acggtcagac tatcaatatt 420
gacctgcagc aaatcgattc tcatacactg ggtctggatg gtttcagcgt taaaaataat 480
gatgcagtga aaaccagtcg tgccgtgaat actcttgggg ggggggcagg ttctgttgct 540
gtcgacttcg caacaaccag tttgactgct atcactgggc tcggtagcgg tgctatcagc 600
gaaattgcta aagacgataa tgggtgattac tacgcgcatg tcacaggggac tacgggtaat 660
actgctgatg gttactatgc tgcgatatc gacaaggcta ccggtgaggt cgctctgaaa 720
gatggtaacg tagatacacc gacaggtagc ccaacgacga caagcacata tgacttcaca 780
gacgctgggc aaaccgtttc ctttggcact gatgctgcaa cagccggtat cagcactggg 840
gcttctctcg ttaaaactca ggatgagaaa ggcaatgata ctgctactta tgcaatcaaa 900
gcacaagatg gcagcctgta tgccgccaac gttgatgagg ctaccggtaa agtcaactgtc 960
aaaaccgcca gctatactga tgctgacggc aaagcagtga ccgatgccgc tgtaaaactg 1020
ggtggtgaca atggcacaac cgaaattgtt gtcgatgctg cgtcaggtaa aacttacgat 1080
gctggtgcac tgcaaaacgt tgatctctcc agtgcaacca acacggtaac cgcaatcccc 1140
aacggtaaaa ccacgtctcc gctggctgcc cttgacgacg caatcagcca gatcgacaaa 1200
ttccgctcct ccctcggtgc ggtgcagaac cgtctggatt ccgcggtcac caacctgaac 1260
aacaccacta ccaacctgtc tgaagcgag tcccgtattc aggacgctga ctatgcgacc 1320
gaagtatcca acatgtcgaa agcgcagatc atccagcagg caggtaactc cgtgctgtcc 1380
aaa
```

<210> 24

<211> 1197

<212> DNA

<213> Escherichia coli

<400> 24

```
gcgctgtcga cttctatcga gcgcctctct tctggtctgc gcattaacag cgctaaagat 60
gacgctgcgg gccaaagcga tgctaaccgc ttcacttcta acatcaaagg tctgactcag 120
gccgcacgta acgccaacga cggatattct ctggcgacga ccaactgaagg cgcactgtct 180
gaaatcaaca acaacttgca gcgtgttcgt gaactgaccg ttcaggccac taccgggtact 240
aactctgatt ctgacctgtc ttcaatacag gacgaaatca aatcccgtct cgatgaaatt 300
gaccgcgtat ccggtcagac tcagttcaac ggcgttaatg ttctttccaa agatggttca 360
atgaaaattc aggttggtgc gaatgatggt caaactatct ccacgatct gaagaaaatt 420
gattcttcaa ctttggggct gaatggcttc tcagtttcta aaaactctct taatgtcagc 480
aatgctatca catctatccc gcaagccgct agcaatgaac ctggtgatgt taacttcggt 540
gatactgatg agtctgcagc aatcgcagcc aaattggggg tttccgatac gtcaagcctg 600
```

```

tcgctgcaca acatccttga taaagatggt aaggcaacag ctgattatgt tgttcagtca 660
ggtaaagact tctatgctgc ttctgttaat gccgcttcag gtaaagtaac cttaaacacc 720
attgatgtta cttatgatga ttatgcgaac ggtgttgacg atgccaagca aacagggtcag 780
ctgatcaaag tttcagcaga taaagacggc gcagctcaag gttttgtcac acttcaaggc 840
aaaaactatt ctgctggtga tgcggcgagc attcttaaga atggagcaac agctcttaag 900
ttaactgata tgaatttaag tgatgttact gatactaata gtaaggtaac cacaactgcg 960
actgagcaat ttgaagggtc ttcaactgag gatccgctgg cgcttctgga taaagctatt 1020
gcatcagtcg acaaattccg gtcttctcta ggtgccgtgc agaaccgtct cgattccgct 1080
atcaccaacc tgaacaacac caccaccaac ctgtctgaag cgcagtcctc tattcaggac 1140
gccgactatg cgaccgaagt gtccaacatg tcgaaagcgc agatcatcca gcaggca 1197

```

<210> 25

<211> 1674

<212> DNA

<213> Escherichia coli

<400> 25

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggccact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctg 360
gacgaaattg acccggtatc tggtcagacc cagttcaacg gcgtgaacgt gctgtctaaa 420
gatggctcga tgaatttca ggtcggcgcg aacgatggcg aaacgattac tattgatctg 480
aagaaaattg actctgatac gctaaatctg gctggtttta acgtgaatgg tgctggctct 540
gttgataatg ccaaggcgac tggcaaagat cttactgatg ctggttttac ggcaagcgca 600
gctgatgcta atggcaaaat cacttatacc aaagacaccg ttactaaatt cgacaaagcg 660
acagcggctg atgtattggg caaagcggct gctggcgata gcattaccta tgcgggcact 720
gatactggct taggagtcgc tgctgatgcc tcgacttaca cctacaatgc agccaataag 780
tcttacactt ttgatgctac tgggtgtgcc aaggcggatg ctggaacggc actgaaaggg 840
tacttaggcg catctaacac cggtaaaatt aatatcgggt gtaccgagca agaagttaac 900
attgccaaag atggtcccat caccgatacc aatggcgatg cgctgtatct cgatagtacc 960
ggcaacttaa ccaaaaatac cgcgaaattg ggggctgctg ataaagcaac tgtagataaa 1020
ctgtttgctg gtgctcagga tgcaacgatc accttcgata gcggcatgac agctaaattc 1080
gatcaactg ctggtaccgt tgatttcaaa ggcgcgtcta tttctgctga tgcaatggca 1140
tcaaccttaa ataattggtc ctatacagcc aacgtagggt gtaaggctta tgccgtaacc 1200
gctggcgagc ttcagacagg tggcgagat gtgtataaag ataccactgg cgcactgacg 1260
actgaagatg acgaaaccgt taccgcgacc tactacgggt ttgctgatgg taaagtttct 1320
gacggtgaag gttctactgt ctataaagct gctgatgggt ccatcactaa agatgcgact 1380
accaagtctg aagcaaccac tgaccctctg aaagcccttg acgacgcaat cagccagatc 1440
gacaaattcc gtcctccct cggtgccgtt caaaaccgtc tggattccgc cgtcaccaac 1500
ctgaacaaca ccactaccaa cctgtctgaa gcgcagtcct gtattcagga cgcogactat 1560
gcgaccgaag tgtccaacat gtcgaaagcg cagatcatc agcaggccgg taactccgtg 1620
ctggcaaaaag ccaaccaggt accgcagcag gttctgtctc tgctgcaggg ttaa 1674

```

<210> 26

<211> 1365

<212> DNA

<213> Escherichia coli

<400> 26

```

aacaaatctc agtcttctct tagctctgct attgagcgctc tctcttctgg cctgcgtatt 60
aacagtgcta aagatgacgc agcaggtcag gcgattgcta accgttttac ggcaaattatt 120
aaaggtctga ctcaggcttc ccgtaacgcg aatgatggta tttctgttgc gcagactact 180
gaaggtgcgc tgaatgaaat taacaacaac ctgcagcgctg tacgtgaact gactgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta ttcaggcaga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gccgaaaata atgaaatgaa aattcagggtt ggtgctaatt atggggaaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggcctggacg gctttaatat cgatggcgcg 480
cagaaagcaa ctggcagtgta cctgatttct aaatttaaag cgacagggtac tgataattat 540
caaattaacg gtactgataa ctatactgtt aatgtagata gtggagcagt tcaaatgag 600
gatggtgacg caatttttgt tagcgctacc gatggttctc tgactactaa gagtataca 660
aaagtcggtg gtacaggtat tgatgcgact gggcttgcaa aagccgcagt ttcttttagct 720
aaagatgcct caattaaata ccaaggtatt actttcacca acaaaggcac tgatgcattt 780
gatggcagtg gtaacggcac tctaaccgct aatattgatg gcaaagatgt aacctttact 840
attgatgcga caggggaagga cgcaacatta aaaacgtctg atcctgttta caaaaatagt 900
gcaggtcagt tctactaac taaggttgaa aacaaagccg ctacagcatc ggatctggac 960
ttaaataacg ctaaaaaagt gggtagttct ttagttgtaa atggcgctga ttatgaagtt 1020
agcgctgatg gtaagacagt aactgggctt ggcaaaaacta tgtatctgag caaatcagaa 1080
ggtggttagc cgattctggt aaaagaagat gcagcaaaat cgttgcaatc tactaccaac 1140
ccgctcgaaa ccctcgacaa ggcattggct aaagttgaca atctgcgttc tgacctcggt 1200
gcagtacaaa accgtttcga ctctgctatc accaaccttg gcaacaccgt aaacaacctg 1260
tcttctgccc gtagccgtat cgaagatgct gactacgcga ccgaagtgtc taacatgtct 1320
cgtgcgcaga tcttgcaaca agcgggtacc tctgttctgg cgcag 1365

```

<210> 27

<211> 1740

<212> DNA

<213> Escherichia coli

<400> 27

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcgcgagg tcaggcgatt gtaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgat ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggt tcaggcttct 300
accgggacta actccgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc tggccagacc cagttcaacg gcgtgaacgt actggcgaaa 420
gacggttcaa tgaaaattca ggttggtgcg aatgacggcc agactatcae gattgatctg 480
aagaaaattg actctgatac gctggggctg agtgggttta atgtgaatgg tagcggggct 540
gtggctaata ctgcagcgac taaatctgat ttggcagcag ctcaactctt ggctccaggt 600
actgctgatg ctaatggtac agttacctat actgttggcg caggcctgaa aacatctaca 660
gctgcagatg taattgcgag tttggctaata aacgcaaaaag ttaatgccac aattgcaa 720
ggttttggat cgccaacagc tacagattat acatacaaca gcgctacagg cgattttaca 780
tatagtgcga ctattgcagc tgggtacaaat tctggtgata gtaacagtgc tcagttacaa 840

```

tccttcctga caccaaaagc gggcgatact gctaacttaa acgttaaaat tggttctacg 900
tcaattgacg ttgtattggc tagcgacggt aaaattaccg cgaaagatgg ttcagaacta 960
tttattgacg tagatggtaa cctcactcaa aacaatgctg ggactgtcaa agcagccact 1020
cttgatgcac tgactaaaaa ctggcataca acaggcacac cgagtgccgt atctacggta 1080
attacaactg aagatgaaac aaccttcact ctggctggcg gtactgatgc tactacttct 1140
gggtgcaatca ctgtagcaaa tgcaagaatg agtgctgagt ctcttcaatc ggcaactaag 1200
tccacaggat tcacagttga tgttggagct actggtacca gcgcaggcga tattaagtt 1260
gatagtaaag gtatagtaca acaacacaca ggtacagggt ttgaagacgc ttacaccaa 1320
gctgatgggt cactgactac cgataatata accaatctgt ttttgcaaaa agacgggaact 1380
gtgaccaatg gttcaggtaa agcagtcctat gtttcagcgg atggtaattt tactactgac 1440
gctgaaacta aagctgcaac caccgccgat ccactgaaag ctctggacga agcgatcagc 1500
tccatcgaca aattccgctt tccctcggt gcggtgcaaa accgtctgga ttccgcagtc 1560
accaacctga acaacaccac tactaacctg tctgaagcgc agtcccgtat tcaggacgct 1620
gactatgcga ccgaagtgtc caatatgtcg aaagcgcaga tcatccagca ggccggtaac 1680
tccgtgctgg caaaagctaa ccaggtagcg cagcagggtt tgtctctgct gcagggttaa 1740

<210> 28

<211> 1233

<212> DNA

<213> Escherichia coli

<400> 28

aacaaaaacc agtctgcgct gtgcacttct atcgagcgcc tctcttctgg tctgcgcatt 60
aacagcgcta aagatgacgc tgccggccag gcgattgcta accgcttcac ttctaacatc 120
aaagggtctga ctcaggccgc acgtaacgcc aacgacggta tctctctggc gcagaccact 180
gaaggcgcac tgtctgaaat caacaacaac ttgcagcgtg ttcgtgagct gaccgttcag 240
gccactaccg gtactaactc tgattctgac ctgtcttcaa tccaggacga aatcaaattc 300
cgtctcgatg aaattgaccg cgtatccggt cagactcagt tcaacggcgt gaacgtactg 360
gcaaaagata acaccatgaa gattcagggt ggtgcgaacg atggtcagac tatatccatc 420
gacctgcaaa aaatcgactc ttctactctt ggtttgaacg gtttctccgt ttctaaaaat 480
gctctcgaaa ctagegaagc gatcactcag ttgccgaacg gtgcgaatgc accaatcgct 540
gtgaagatgg atgcgtctgt tctgaccgat cttaacatta ctgatgcttc cgctgtttcg 600
ctgcacaacg taactaaagg tgggtgogca acgtctactt atgttggtca gtatggcgat 660
aagagctatg cagcatctgt tgatgcggga ggtacagtaa aactgaataa agccgacgta 720
acataatacg acgcagcaaa tgggtgttacg aatgccaccc agattggtag tctgggttcag 780
gttgggtgctg atgcaaaca tgcagcagtt ggttttgta ccgtagcagg gaaaaactat 840
gttgctaatt actcattagt caatgctaatt ggcgctgctg gcgctgcagc aactagagtt 900
acaattgatg gtgatggtag ccttggagct aaccaggcta aaattgaact tagccaaaat 960
gggtgctactg ctgcaacatc agagtctgct ggtgcttcaa ccaacgatcc actgactctg 1020
ctggacaaag ctatcgcatc tggtgataaa ttccgttctt ctttgggggc ggtacagaac 1080
cgtctgagct ccgctgtaac caacctgaac aacaccacta ccaacctgtc tgaagcgag 1140
tcccgatttc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1200
atccagcagg caggtaactc cgtgctgtcc aaa 1233

<210> 29

<211> 1713

<212> DNA

<213> Escherichia coli

- 19 -

<400> 29

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttactttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggcgacg 300
accggaacta actccacctc tgacctggac tccattcagg acgaaatcaa atcccgtctt 360
gatgaaattg accgcgtatc cggccaaacc cagttcaacg gcgtgaacgt actgtcaaaa 420
gatggctcga tgaaaattca ggtcggcgca aatgatggtg aaaccatcac gattgatctg 480
aaaaagatcg actcttctac attgaagctg accagcttca atgttaacgg taaaggcgct 540
gttgataatg ctaaagccac tgaagcagat ctgaccgctg cgggcttctc ccaagggtgca 600
gtcgctcagt gcaacagcac ctggactaaa tctactgtta ctacctttaa tgcagcaaca 660
gctaccgacg tgctggcaag cgtagcggc ggcagcacta ttagcgggta taccggtaca 720
aacaatggat taggcgtagc ggcttctact gcatatacct acaacgcaac cagcaagtct 780
tattcatttg acgcaaccgc acttaccaat ggcgatggta ctggggccac cactaaagtt 840
gctgatgtgc tgaaagccta tgcagcaaac ggtgataata cggctcagat ctccatcggc 900
ggaagcgctc aggacgttaa aattgccagc gatggcacc tgactgacgt caatggtgat 960
gctttatata ttggttctga cggcaacctg actaaaaacc aggccggcgg tccagatgcg 1020
gcaacggttg acggtatttt caacggtgcg aatggtaatg cagcagttga tgcgaagatt 1080
acattcggca gcggcatgac cgttgatttc accaggcta gcaaaaaagt ggatattaag 1140
ggcgcaacgg tatccgccga agatatggac actgcgttaa ctgggcaggc ttataccgta 1200
gctaaccggc cacagtcttt tgacgttgcc gctgggtggg cagtaaccgc tactacaggt 1260
ggcgctaccg taaatatttg tgctgatggt gaactgacga ctgcgaccaa caagactgtc 1320
acagaaaact atcacgaatt tgctaacggc aatattctgg atgatgacgg cgcggtctctg 1380
tacaaagcgg ctgacgggtc tctgaccact gaagctactg gtaaattccga agtgaccacg 1440
gatccgctga aagcgctgga cgatgctatc gcatccgtag acaaattccg ctctccctc 1500
ggtgcggtgc agaaccgtct ggattccgca gtcaccaacc tgaacaacac cactaccaac 1560
ctgtctgaag cgcagtcccg cattcaggac gccgactatg cgaccgaagt gtccaatatg 1620
tcgaaagcgc agatcatcca gcaggccggt aactccgtgc tggcaaaagc caaccaggta 1680
ccgcagcagg ttctgtctct gctgcagggt taa 1713

```

<210> 30

<211> 1668

<212> DNA

<213> Escherichia coli

<400> 30

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gctaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttactttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggt tcaggcttct 300
accgggacta actccgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc tggccagacc cagttcaacg gcgtgaacgt actggcgaaa 420
gacggttcaa tgaaaattca ggttggtgcg aatgacggcc agactatcac tattgatctg 480
aagaaaattg actcagatac gctggggctg agtgggttta atgtgaatgg tggcggggct 540
gttgctaata ctgcagcgac taaagatgat ttggctgctg catcagtttc agctgcggta 600

```

```

ggtaaatgaat acactgtctc tgctggcctg tcgaaatcaa ctgctgctga tgttattgct 660
agtctcacag atggtgcgac agtaactgcg gctggtgtaa gcaatgggtt tgctgcaggg 720
gcaactggag atgcttataa attcaatcaa gcaacaaca cttttactta caataccacc 780
tcaacagcgg cagaactcca atcttacctc acgcctaagg cgggggatac cgcaactttc 840
tccgttgaaa ttggtggcac caagcaggat gttgttctgg ctagtgatgg caaaatcaca 900
gcaaaagacg ggtctaaact ttatattgac accacagga atttaacca aaacggtgga 960
ggtacttttag aagaagctac cctcaatggc ttagctttca accactctgg tccagccgct 1020
gctgtacaat ctactattac tactgcggtt ggaacttcaa tagttctagc aggttctggc 1080
gacttttgaa caacaaaaac tgctggggct attaattgca caggagcagt gatcagtgtc 1140
gatgcacttc tttccgccag taaagcgact gggtttactt ctggcactta taccgtaggt 1200
acagatggag ttgttaaatc tgggtggcaat gacgtttata acaaagctga cgggacggga 1260
ttaactactg acaataccac aaaatattat ttacaagatg acgggtctgt aactaatggt 1320
tctggtaaag ctgtgtatgc tgatgcaaca ggaaaactaa ctactgacgc tgaaactaaa 1380
gccgaaacca ccgccgatcc cctgaaagct ctggacgaag cgatcagctc catcgacaaa 1440
ttccgttctt ccctcgggtg ggtgcaaaac cgtctggatt ccgcggtcac caacctgaac 1500
aacaccacta ccaacctgtc cgaagcgcag tcccgtattc aggacgccga ctatgcgacc 1560
gaagtgtcca acatgtcgaa agcgcagatc atccagcagg ccggttaactc cgtgctggca 1620
aaagctaacc aggtaccgca gcaggttctg tctctgctgc aggggttaa 1668

```

<210> 31

<211> 1713

<212> DNA

<213> Escherichia coli

<400> 31

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttccg ttgcgcagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggt tcaggccact 300
accggtacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
gatgaaattg accgcgtatc tggtcagacc cagttcaatg gcgtgaatgt gttgtccaaa 420
gacggttcaa tgaaaattca ggtgggcgca aatgatgggt aaaccatcac gattgacctg 480
aaaaaaaaatc actcttctac actgaagctg accagettca acgtcaacgg taaaggcgct 540
gttgataatg caaaagccac tgaagcagat ctgaccgctg cgggcttctc ccaaagtga 600
gttgctcagt gcaatagcac ctggactaaa tctactgtta ctacctttaa tgcagcaaca 660
gctaccgatg tgctggctag cgttagtggc ggcagcacta ttagcgggta tgctggcaca 720
aacaatgggt taggcgtagc ggcttctact gcatatacct acaacgcaac cagcaagtct 780
tattcatttg acgcaaccgc acttactaat ggtgatggta ctgcgggctc aactaaagtt 840
gctgatgttc tgaaagccta tgcagcaaac ggcgataaca cggctcagat ctccatcggt 900
ggtagcgctc aggaagttaa aattgccagc gatggtaccc tgacgggatac taatggcgat 960
gctttataca ttggtgctga cggtaacctg acgaaaaacc aggccggcgg cccagcccg 1020
gcaacgttg acggtatttt caacggtgcg aatggtcatg atgcagttga tgcgaagatt 1080
accttcggca gcggcatgac cgttgacttc acccagggtt gcaacaatgt ggatattaag 1140
ggcgcgacgg tatccgccga agatatgaac actgcgttaa ccggtcaggc ttataccgta 1200
gctaacggcg cacagtctta tgacgttgcc gctgatgggt cagtaactgc tactacaggt 1260
ggagcgaccg taaatattgg tgctgagggt gaactgacga ctgcggccaa caagactgtc 1320
acagaaactt atcacgaatt tgctaacggc aatattctgg atgatgacgg cgcggtctct 1380

```

tataaagcgg ctgacggctc tctgaccact gaagctacag gtaaactctga agcgaccacg 1440
 gatccgctga aagcgctgga cgatgctatc gcatccgtag acaaattccg ttcttccctg 1500
 ggtgcccgtgc agaaccgtct ggattccgca gtcaccaacc tgaacaacac cactaccaac 1560
 ctgtccgaag cgcagtcccc tattcaggac gccgactatg cgaccgaagt gtccaacatg 1620
 tcgaaagcgc agattattca gcaggcaggt aactccgtgc tggcaaaagc taaccaggta 1680
 ccgcagcagg ttctgtctct gctgcagggt taa 1713

<210> 32

<211> 1188

<212> DNA

<213> Escherichia coli

<400> 32

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgcatt 60
 aacagcgcta aagatgacgc tgcgggccag gcgattgcta accgcttcac ttctaacatc 120
 aaaggctctga ctgagccgc acgtaacgcc aacgacggta tctctctggc gcgaccact 180
 gaaggcgac tgtctgaaat caacaacaac ttgcagcggtg tgcgtgagtt gactgttcag 240
 gcgacgaccg ggactaactc tgattctgac ctgtcttcta ttcaggacga aatcaaattc 300
 cgtctggatg aaattgaccg tgtttccggt cagaccaggt tcaacggcgt gaacgtgctg 360
 gctaaaaacg gttctatggc gattcagggt ggcgcgaaatg atgggcagac catcaacatc 420
 gacctgcaga aaatcgactc ttctactctg ggcttggcg gcttctccgt atctaacaat 480
 gcaactgaaac tgagcgattc tatcactcag gttgggtgca gtgggttact ggcagatgtg 540
 aaactgagct ctgttgctc ggctctgggt gtagacgcaa gcactctgac tctgcacaac 600
 gtacagaccc cagctggcgc agcaacagct aactatgttg tctcttctgg ttctgacaac 660
 tactcagtat ctgttgaaga tagctccggt acagttacgc tgaacaccac tgatataggt 720
 tataccgata ccgctaattg cgttactacc ggttccatga ctggtaagta cgttaaagtt 780
 ggagctgatg cattgggtgc tgctgtaggt tatgtcaccg tacagggaca aaacttcaaa 840
 gctgatgctg gcgcgctggt taactccaag aatgctgctg gtagtcagaa tgttacttct 900
 gcaattggcg atattgctaa taaagcgaat gctaacattt aacttggaaac ctcttctgca 960
 gatccactgg ctctgctgga caaagctatc gcatctgttg ataaattccg ttcttctcta 1020
 gggcggtgc agaaccgtct gagctctgct gtaaccaacc tgaacaacac cactaccaac 1080
 ctgtccgaag cgcagtcccc tattcaggac gccgactatg cgaccgaagt gtccaacatg 1140
 tcgaaagcgc agatcatcca gcaggcgggt aactccgtgc tgtctaaa 1188

<210> 33

<211> 1638

<212> DNA

<213> Escherichia coli

<400> 33

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaataa tatcaacaag 60
 aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
 gcgaaggatg acgccgcggg tcaggcgatt gctaaccggt ttacttctaa cattaaaggc 180
 ctgactcagg ctgcacgtaa cgccaatgac ggtatttctg ttgcacagac cactgaaggc 240
 gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggcttct 300
 accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
 gacgaaattg accgcgtatc cggctcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
 gacggttcga tgaaaattca gggtggtgcg aacgacggcc agactatcac tattgatctg 480

aagaaaattg actctgatac gctggggctg agtgggttta acgtaaatgg tagcgagat 540
aaggcaagtg tcgcggcgac agctgacgga atgggttaaag acggatatat caaaggggta 600
acttcatctg acggcagcac tgcataact aaaactacag caaatactgc agcaaaagga 660
tctgatattc ttgcggcgct taagactggc gataaaatta ccgcaacagg tgcaaatagc 720
cttgctgata atgcgacatc gacaacttat acttataatg caaccagcaa taccttctcc 780
tatacggctg acgggtgtaaa ccaaacgaat gctgcagcaa atctcatacc tgcagcaggg 840
aaaacgacag ctgcatcagt tactattggg gggacagcac agaattgtaaa tattgatgat 900
tcgggcaata ttacttcaag tgatggcgat caactttatc tggattcaac aggtaacctg 960
actaaaaacc aggcgggcaa cccgaaaaaa gcaaccgttt ctgggcttct cggaataacg 1020
gatgcgaaag gtactgctgt taaaacaacc atcaagacag aggctggtgt aacagttaca 1080
gctgaaggta atacaggtac tgtaaaaatt gaagggtgcta ctgtttcagc atctgcattt 1140
acgggcattg catattccgc caacaccggt gggaataact atgctgttgc cgcaataat 1200
actacaaatg gtttctctggc gggggatgac ttaaccagg atgctcaaac tgtttcaacc 1260
tactactcgc aagccgatgg caggttcacg aatagcgagc gcaaagaaat ctataagac 1320
gctgatggtg tctacagcac agagaataaa acatcgaaga cgtccgatcc attggctgcg 1380
cttgacgacg caatcagctc catcgacaaa ttccgttcat ccttgggtgc tatccagaac 1440
cgtctggatt ccgcggtcac caacctgaac aacaccacta ccaacctgtc cgaagcgagc 1500
tcccgtatcc aggcgcgga ctatgcgacc gaagtgtcca acatgtcgaa agcgagatc 1560
atccagcagg ccggttaact cgtgctggca aaagctaacc aggtaccgca gcaggttctg 1620
tctctgctgc agggctaa 1638

<210> 34

<211> 2145

<212> DNA

<213> Escherichia coli

<400> 34

aacaaatctc agtcttctct gagctccgcc attgaacgtc tctcttctgg cctgcgtatt 60
aacagtgcta aagatgacgc agcaggtcag gcgattgcta accgttttac agcaaatatt 120
aaagggtctga ctcaggcttc ccgtaacgcg aatgatggtg tttctgttgc gcagaccact 180
gaagggtgcgc tgaatgaaat taacaacaac ctgcagcgtg tacgtgaact gactgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggtt ggtgctaatt atggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggccctggacg gttttaatat cgatggcgcg 480
cagaaagcaa ctggcagtga cctgatttct aaattttaaag cgacaggtac tgataactat 540
gatgttggcg gtgatgctta tactgttaac gtagatagcg gagctgggta atgactccaa 600
cttattgata gtgttttatg ttcagataat gccgatgac tttgtcatgc agctccaccg 660
attttgagaa cgacagcgac ttccgtccca gccgtgccag gtgctgcctc agattcagg 720
tatgccgctc aattcgtgc gtatategt tgctgattac gtgcagctt cccttcaggc 780
gggattcata cagcggccag ccacccgtca tccatatcac cagctcaaag ggtgacagca 840
ggctcataag acgcccagc gtcgccatag tgcgttcacc gaatacgtgc gcaacaaccg 900
tcttccggag cctgtcatac gcgtaaaaca gccagcgtg gcgcgattta gccccgacat 960
agtcccactg ttcgtccatt tccgcgcaga cgatgacgtc actgcccggc tgtatgcgcg 1020
aggttaccga ctgcggcctg agttttttaa gtgacgtaaa atcgtgttga ggccaacgcc 1080
cataatgcgg gcagttgccc ggcattccaac gccattcatg gccatatcaa tgattttctg 1140
gtgcgtaccg ggttgagaag cgggtgaagt gaactgcagt tgccatgttt tacggcagtg 1200
agagcagaga tagcgtgat gtccggcggt gcttttgccg ttacgcacca ccccgctcag 1260

agctgaacag gagggacagc tgatagaaac agaagccact ggagcacctc aaaaacacca 1320
tcatacacta aatcagtaag ttggcagcat taccgcggag ctgttaaaga tactacaggg 1380
aatgatattt ttgttagtgc agcagatggt tctactgacaa ctaaactctga cacaacata 1440
gctggtacag ggattgatgc tacagcactc gcagcagcgg ctaagaataa agcacagaat 1500
gataaattca cgtttaatgg agttgaattc acaacaacaa ctgcagcggg tggcaatggg 1560
aatggtgtat attctgcaga aattgatggt aagtcagtga catttactgt gacagatgct 1620
gacaaaaaag cttctttgat tacgagttag acagtttaca aaaatagcgc tggcctttat 1680
acgacaacca aagttgataa caaggctgcc acactttccg atcttgatct caatgcagct 1740
aagaaaacag gaagcacgtt agttgttaac ggtgcaactt acgatgttag tgcagatggt 1800
aaaacgataa cggagactgc ttctggtaac aataaagtca tgtatctgag caaatcagaa 1860
ggtggttagc cgattctggt aaacgaagat gcagcaaaat cgttgcaatc taccaccaac 1920
ccgctcgaaa ctatcgacaa agcattggct aaagttgaca atctgcgttc tgacctcggt 1980
gcagtacaaa accgtttcga ctctgctatc accaaccttg gcaacaccgt aaacaacctg 2040
tcttctgccc gtagccgtat cgaagatgct gactacgcga ccgaagtgtc taacatgtct 2100
cgtgcgcaga tcctgcaaca agcgggtacc tctgttctgg cgcag 2145

<210> 35

<211> 1587

<212> DNA

<213> Escherichia coli

<400> 35

aacaagaacc agtctgcgct gtcgagttct atcgagcgtc tgtcttctgg cttgcgtatt 60
aacagcgcga aggatgacgc cgcaggtcag gcgattgcta accgttttac ttctaacatt 120
aaaggcctga ctcaggctgc acgtaacgcc aacgacggta tttctgttgc gcagaccacc 180
gaaggcgcgc tgtccgaaat caacaacaac ttacagcgtg tgcgatgaact gaccgttcag 240
gcaaccaccg gtaccaactc ccagttctgac ctggactcta tccaggacga aattaaatcc 300
cgtctggacg aaattgaccg cgtatccggt cagaccagct tcaacggcgt gaacgtactg 360
gcaaaagacg gttccatgaa aattcagggtt ggcgcgaacg atggccagac catcactatc 420
gacctgaaga agattgactc ttctacgctg aaactgactg gttttaacgt gaatggcaaa 480
gcagcgggtg ataattgtaa agcgacggat gcaaatctga ctaccgcggg ttttacacaa 540
ggcgttggtg attcaaatgg taatagtact tggactaaat caactacgac taatttcgat 600
gcggcaactg cagtaaacgt actagcagca gttaaagatg gcagcacaat caattacacc 660
ggtactggta atggttttagg gattgctgca acaagtgtt atacatatca cgatagcact 720
aaatcctata cttttgatcc tacgggggct gcagtagctg gtgccgcgtc cagcctgcaa 780
ggtacttttg gtacagatac gaatactgca aaaatcacca tcgatgggtc tgctcaagaa 840
gtaaacatcg ctaaagatgg gaaaattact gatactgatg gttaaagctt atatatcgat 900
tccactggta atttgactaa gaacggctct gatactttaa ctcaggcaac attgaatgat 960
gtccttactg gtgctaattc agttgatgat acaaggattg acttcgatag cggcatgtct 1020
gtcacccttg ataaagtga cagcactgta gatatcactg gcgcatctat ttcagccgct 1080
gcaatgacta atgagttgac aggtaaggcc tataccgtag taaatgggtc agaactctac 1140
gctgtagcta ctaataacac agtaaaaacg actgctgatg ctaaaaatgt ttatgttgat 1200
gctagtggta aattaactac tgatgacaaa gccactgtta cagaaactta tcatgaattt 1260
gcgaatggca atatctatga tgataaaggc gctgctgttt atgcggcggc ggatgggttct 1320
ctgactacag aaactacaag taaatcagaa gctacagcta acccgctggc cgctctggac 1380
gacgcaatca gccagatcga caaatccgt tcacccctgg gtgctatcca gaaccgtctg 1440
gattccgcag tcaccaacct gaacaacacc actaccaatc tgtctgaagc gcagtccctg 1500
attcaggacg ccgactatgc gaccgaagtg tccaatatgt cgaaagcgca gatcatccag 1560

caggcaggca actccgtgct ggcaaaa

1587

<210> 36

<211> 1245

<212> DNA

<213> Escherichia coli

<400> 36

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgcat 60
aacagcgcta aagatgacgc tgcgggccag gcgattgcta accgcttcac ttctaacatc 120
aaaggctctga ctcaggccgc acgtaacgcc aacgacggta tctctctggc gcagaccact 180
gaaggcgcac tgtctgaaat caacaacaac ttgcagcggtg ttcgtgaact gaccgttcag 240
gccactaccg gtactaactc tgattctgac ctgtcttcaa tccaggacga aatcaaattcc 300
cgtctcgatg aaattgaccg cgtatccggt cagactcagt tcaacggcgt gaacgtactg 360
gcaaaagatg gctcgatgaa aattcaggtc ggtgcaaatg atggtcagac aatcagcatt 420
gatttgcaga agattgattc ttctacttta ggggttaaatg gtttttctgt ttccaaaaat 480
gcagtatctg ttggtgatgc tattactcaa ttgcctggcg agacggcagc cgatgcacca 540
gtaaccatca agtttgatga ttcagtaaaa actgatttaa aactgaccga tgcttcaggg 600
ttaagtctgc ataacctcaa agatgaaaat ggtaatttaa ctaaccagta tgttgtagag 660
aatggcgga aatcttacgc tgctacagtc gctgccaatg gtaatgttac gctgaacaaa 720
gcaaatgtaa cctacagcga tgcgcaaac ggtattgata ccgcaacgca gtcaggccag 780
ttagttcagg ttggtgcaga ttctaccggt acgcaaaaag cattcgtgtc tgtccaaggt 840
aaaagctttg gcattgatga cgcgccttg aagaataaca ctgggtgatgc taccgctact 900
caaccgggaa catctgggac aacagttgtc gcagcgtaaa ttcattctgag tacgggcaaa 960
aactctgtag acgctgatgt aacggcttcc actgaattca cagggtgctt acccaacgat 1020
ccactgactc tgctggacaa agctatcgca tctgttgata aattccgttc ttctttgggg 1080
gcggtacaga accgtctgag ctccgctgta accaacctga acaacaccac caccaacctg 1140
tctgaagcgc agtcccgat tccaggacgc gactatgcga ccgaagtgtc caacatgtcg 1200
aaagcgcaga ttatccagca ggcaggtaac tccgtgctgt ccaaa 1245

<210> 37

<211> 1185

<212> DNA

<213> Escherichia coli

<400> 37

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgcat 60
aacagcgcta aagatgacgc tgcgggccag gcgattgcta accgcttcac ttctaacatc 120
aaaggctctga ctcaggctgc acgtaacgcc aatgacggta tttctctagc acagacagcg 180
gaaggcgcg cgtcagagat taacaacaac ttgcagcggtg tgcgtgagtt gaccgtgcag 240
gcaaccactg gtaccaactc tgattccgat ctctcttcta ttcaggatga aattaaatct 300
cgtctggatg aaattgaccg cgtctctggt cagaccagc ttaacggcgt gaacgtactg 360
gctaaaaacg gttctatggc aattcagggtt ggcgcgaacg atggccagac tatctctatc 420
gacctgcaga aatagactc ttctactctg ggtctgagcg gcttctctgt ttctcagaac 480
tccctgaaac tgagcgattc tatcactacg atcggcaata ctactgctgc atcgaagaac 540
gtggacactg gcgcagtagc aactaaactg ggcgtgaatg caagcaccct gagcctgcac 600
gaagttcagg actctgctgg tgacggtact ggtaccttcg ttgtttcttc tggcagcgac 660
aactatgctg tgtctgtaga cgcggcctct ggtgcagtta acctgaacac cactgacgtc 720

acctatgatg acgctactaa tgggtgttact ggcgcgactc agaacgggtca gctgatcaaa 780
gtaacttctg acgccaacgg tgcagctgtt ggttacgtaa ccattcaggg taaaaactat 840
caggctgggtg cgaccgggtg tgacgttctg gcgaacagcg gtgttgccagc tccaactaca 900
gctgttgata cgggtactct gcaactgagc ggtactgggtg caactactga gctgaaaggt 960
actgcaactc agaaccactc ggcactattg gacaaagcta tcgcttctgt tgataaattc 1020
cgttcttctc tgggtgacgg acagaatcgt ctgagctctg ctgtaaccaa cctgaataac 1080
accaccacta acctgtctga agcgcagtcg cgtattcagg atgccgacta tgcgaccgaa 1140
gtgtcaaata tgtctaaagc gcagatcggt cagcaggccg gtaac 1185

<210> 38

<211> 1383

<212> DNA

<213> Escherichia coli

<400> 38

aacaaatctc agtcttctct tagctctgct attgagcgtc tgtcttctgg tctgcgtatt 60
aacagcgcaa aagacgatgc agcaggtcag gcgattgcta accgttttac ggcaaatatt 120
aaaggtctga ccagagcttc ccgtaacgca aatgatggta tttctgttgc gcagaccact 180
gaaggtgcgc tgaatgaaat taacaacaac ctgcagcgta ttcgtgaact ttctgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggtt ggtgctaata atggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggccctggacg gttttaatat cgatggcgcg 480
cagaaagcaa caggcagtgat cctgatttct aaatttaag cgacaggtag tgataattat 540
gatgttggcg gtaaaactta taccgtgaat gtggagagcg gcgcggttaa gaatgatgct 600
aataaagatg tttttgtaag cgcagctgat ggatcgctga cgaccagtag tgatactaaa 660
gtatccggtg aaagtattga tgcaacagaa ctacgcaaac ttgcaataaa attagctgac 720
aaaggctcca ttgaatacaa gggcattaca tttactaaca aactggcgcg agagcttgat 780
gctaattggt aaggtgtttt gaccgcaaatt attgatggtc aagatgttca atttactatt 840
gacagtaatg caccacaggg tgccggcgca acaataacta cagacacagc tgtttacaaa 900
aacagtgcgg gccagttcac cactacaaaa gtggaaaata aagccgcaac actctctgat 960
ctggatctta atgcagccaa gaaaacaggt agcacttttag ttgtaaatgg cgccacctac 1020
aatgtcagcg cagatggtaa aacggtaact gatactactc ctgggtgccc taaagtgatg 1080
tatctgagca aatcagaagg tggtagcccc attctggtaa acgaagatgc agcaaaatcg 1140
ttgcaatcta ccaccaacct gctcgaaact atcgacaagg cattggctaa agttgacaat 1200
ctgcgttctg acctcggtgc agtacaaaac cgtttcgact ctgccatcac caaccttggc 1260
aacaccgtaa acaacctgtc ttctgcccgt agccgtatcg aagatgctga ctacgcgacc 1320
gaagtgtcta acatgtctcg tgcgcagatc ctgcaacaag cgggtacctc tgttctggcg 1380
cag 1383

<210> 39

<211> 1680

<212> DNA

<213> Escherichia coli

<400> 39

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120

gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt tcacctctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggt tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacggttcaa tgaaaattca ggttggtgcg aatgacggcc agactatcac tattgatctg 480
aagaaaattg actctgatac tctgggtttg agtggattta atgtgaatgg caaaggggct 540
gtggctaacg caaaagcgac cgaagcagat ttaacggggg ctggtttctc tcaaggagcg 600
gtggatacaa acggaaatag tacttgga caaatcaacca ccaccaatta ctcagctgca 660
acaactgctg acttggtatc gaccattaag gatggctcta ctgttacata tgcagggaca 720
gacaccggat taggggtcgc agcagcagga aattatactt atgatgcgaa cagtaaactc 780
tattccttca atgccaatgg tctgacgggc gcaaataaccg caactgcact caaagggttac 840
ttggggacag gtgctaacac cgctaaaatt tctatcggtg gtacagagca ggaagtgaat 900
attgccaag atggcactat tacagatacg aatgggtgat cgctctatct ggatattacc 960
ggcaacctga ctaagaacta tgcgggttca ccacctgcag caacgctgga taacgtatta 1020
gcttccgcaa ctgtaaatgc cactatcaag tttgatagcg gtatgacggt tgattacact 1080
gcagggtactg gcgcgaatat tacagggtgca tccatttctg cagatgacat ggccgcaaaa 1140
ctgagcggaa aggcgtacac tgttgccaat ggtgctgagt cttatgacgt tgctgcagtt 1200
acgggggctg taacaactac agcaggtaat tcacctgtgt atgccgatgc agacggtaaa 1260
ttaacgacga gtgccagtaa tacgggtact cagacttacc acgagtttgc taatggtaac 1320
atztatgatg acaaaggctc gtcactgtat aaagctgcag atggctctct gacttctgaa 1380
gctaaagggg aatctgaagc aaccgccgat cccctgaaag ctctggacga agccatcagc 1440
tccatcgaca aattccgctc ctccctcggt gccgttcaaa accgtctgga ttctgcgggtg 1500
accaacctga acaacaccac taccaacctg tctgaagcgc agtcccgtat tcaggacgcc 1560
gactatgcga ccgaagtgtc caatatgtcg aaagcgcaga tcaccagca ggccggtaac 1620
tccgtgttgg caaaagctaa ccaggtagcg cagcagggtc tgtctctgct gcagggttaa 1680

<210> 40

<211> 1146

<212> DNA

<213> Escherichia coli

<400> 40

gcgctgtcga cttctatcga gcgcctctct tctggtttgc gcattaacag cgctaaagat 60
gacgctgcgg gccaggcgat tgctaaccgc ttcacttcta acatcaaagg tctgactcag 120
gccgcacgta acgccaacga cggatctctc ctggcgcaga cactgaagg cgcactgtct 180
gaaatcaaca acaacttgca gcgtgttcgt gaactgaccg ttcaggccac taccggtact 240
aactctgatt ctgacctgtc ttcaatccag gacgaaatca aatcccgtt ggctgaaatc 300
gatcgtgtct ctggtcagac ccagttcaac ggctgaacg tgctggctaa aaacggttct 360
ctgaatattc aggttggcgc gaatgatggg cagaccatct ctatcgattt gcagaaaata 420
gactcttctg cccttggttt aagtggtttt agtgttgccg gtggggcgct aaaattaagc 480
gatacagtga cgcaggtcgg cgatggttca gccgcgccag ttaaagtgga tctggatgca 540
gcagcaacag atattggtac tgctttgggg caaaagggtta atgcaagttc tttaacggtg 600
cacaatatct tagacaaaaga tgggtgcggca actgagaact atgttggttag ctatggtagt 660
gataattacg ctgcatctgt tgcagatgac gggactgtaa ctcttaataa aacggatatt 720
acttattcag gcggtgatat taccggcgct accaaagatg atacgttgat taaagttgct 780
gctaattctg acggagaggg cggttggtttc gctaccgttc agggtaagaa ttatgaaatt 840
acagatgggtg taaaaaacca gtccactgct gcaccaaccg atattgctca gaccattgat 900

ctggatacgg ctgatgaatt tactggggct tccactgctg atccactggc acttttagac 960
aaagctattg cacaggttga tactttccgc tcttccctcg gtgccgttca aaaccgtctg 1020
gattccgcag tcaccaacct gaacaacact actaccaacc tgtctgaagc gcagtcacct 1080
attcaggacg ccgactatgc gaccgaagtg tccaatatgt cgaaagcgca gatcatccag 1140
caggcc 1146

<210> 41

<211> 1506

<212> DNA

<213> Escherichia coli

<400> 41

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccgtt ttacttctaa tattaaaggc 180
ctgactcagg ctgcacgtaa cgccaatgac ggtatttctc tggcgcagac cactgaaggc 240
gcactgtctg aaatcaacaa caacttgacg cgtgtgcgtg aactgaccgt acaggcgaca 300
accggaacga actccgaatc tgacctgtcc tctatccagg acgaaatcaa atcccgctctg 360
gaagagattg accgcgtatc cggccagact cagttcaacg gcgtgaatgt gctggcaaaa 420
gacggcacca tgaaaattca ggtaggcgag aacgatggtc agactatctc tatcgatctg 480
aaaaaaatcg actcttcaac cctgggcctg accggttttg atgtttcgac gaaagcgaat 540
atttctacga cagcagtaac gggggcggca acgaccactt atgctgatag cgccgttgca 600
attgatatcg gaacggatat tagcggattt gctgctgatg ctgcgttagg aacgatcaat 660
ttcgataata caacaggcaa gtactacgca cagattacca gtgcggccaa tccgggcctt 720
gatggtgctt atgaaatcca tgtaatgac gcggatgggt ccttcaactgt agcagcgagt 780
gataaacaag cgggtgctgc tccgggtact gctctgacaa gcggtaaagt tcagactgca 840
accaccacgc caggtagcgc tgttgatgtc actgcggcta aaactgctct ggctgcagca 900
ggtgctgaca cgagtggcct gaaactgggt caactgtcca acacggattc cgcaggtaaa 960
gtgaccaacg tgggttacgg cctgcagaat gacagcggca ctatctttgc aaccgactac 1020
gatggcacca ctgtgaccac gccgggcgca gagactgtga cttacaaaga tgcttccggt 1080
aacagcacca ctgcggctgt cacactgggt ggctctgatg gcaaaaccaa tctggttacc 1140
gccgctgacg gcaaaacgta cgggtgcgact gcaactgaatg gtgctgatct gtccgatcct 1200
aataacaccg ttaaactctgt tgcagacaac gctaaaccgt tggctgccct ggatgatgca 1260
attgcgatgg tcgacaaaatt ccgtctctcc ctccggtgcg tgcaaaaccg tctggattcc 1320
gcagtcacca acctgaacaa caccactacc aacctgtctg aagcgcagtc ccgtattcag 1380
gacgccgact atgcgaccga agtgtccaac atgtcgaaag cgcagattat ccagcaggca 1440
ggtaactccg tgctgtccaa agctaaccag gttccgcagc aggttctgtc tctgctgcag 1500
ggttaa 1506

<210> 42

<211> 950

<212> DNA

<213> Escherichia coli

<400> 42

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgtatt 60
aacagcgcta aagatgacgc cgcggggccag gcgattgcta accgctttac ttctaacatc 120
aaaggtctga ctcaggccgc acgtaacgcc aacgacggta tttctctggc gcagacggct 180

```

gaaggcgcgc tgtcagagat taacaacaac ttgcagcgta ttcgtgaact gaccgttcag 240
gcctctaccg gcacgaactc tgattccgac ctgtcttcta ttcaggacga aatcaaatec 300
cgtcttgatg aaattgaccg tgtatctggc cagaccagc tcaacgggtg gaacgtgctg 360
tcgaaaaacg attcagatgaa gattcagatt ggtgccaatg ataaccagac gatcagcatt 420
ggcttgcaac aaatcgacag taccactttg aatctgaaag gatttaccgt gtccggcatg 480
gcggtattca gcgcggcgaa actgacggct gctgatggta cagcaattgc tgctgcggtg 540
gtcaaggatg ctgggggtaa acaagtcaat ttactgtctt acactgacac cgctgtctaac 600
agtactaaat atgcggtcgt tgattctgca accggtaaat acatggaagc cactgtagcc 660
attaccggta cggcggcggc ggtaactgtt ggtgcagcgg aagtggcggg agccgctaca 720
gccgatccgt taaaagcact ggatgccgca atcgctaaag tcgacaaatt ccgtctctcc 780
ctcgggtgcc ttcaaaaccg tctggattct gcggtcacca acctgaacaa caccaccacc 840
aacctgtctg aagcgcagtc ccgtattcag gacgccgact atgcgaccga agtgtccaac 900
atgtcgaaag cgcagattat ccagcaggcc ggtaactccg tgctggcaaa 950

```

<210> 43

<211> 1707

<212> DNA

<213> Escherichia coli

<400> 43

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccgtt ttacctctaa cattaaaggt 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg aactgacggg tcaggcttct 300
accgggacta actccgattc ggatctggac tccattcagg acgaaatcaa atcccgctctg 360
gacgaaattg accgcgtatc cgggtcaaac cagttcaacg gtgtgaacgt actggcgaaa 420
gacggttcga tgaaaattca ggttggtgcg aatgacggcc agactatcac gattgatctg 480
aagaaaattg actcagatac gctggggctg aatgggttca acgttaatgg caaaggcact 540
attgcgaaca aagctgctac agtcagcgat ctgaccgctg ctggtgcaac gggaacaggt 600
ccttatgctg tgaccacaaa caatacagca ctacgcgcta gcgatgact gtctcgccctg 660
aaaaccggag atacagttac tactactggc tcgagtgtg cgatctatac ttatgatgcg 720
gctaaaggga acttcaccac tcaagcaaca gttgcagatg gcgatgttgt taactttgcg 780
aatactctga aaccagcggc tggcactact gcacaggtg tttatactcg tagtactggt 840
gatgtgaagt ttgatgtaga tgctaattgg gatgtgacca tcggttggtaa agcccggtac 900
ctggacgcca ctggtaacct atctacaaac aaccccgcca ttgcatcttc agcgaaattg 960
tccgatctgt ttgctagcgg tagtacctta gcgacaactg gttctatcca gctgtctggc 1020
acaacttata actttggtgc agcggcaact tctggcgtaa cctacaccaa aactgtaagc 1080
gctgatactg tactgagcac agtgcagagt gctgcaacgg ctaacacagc agttactggt 1140
gcgacaatta agtataatac aggtattcag tctgcaacgg cgtccttcgg tgggtgtaat 1200
actaatggtg ctggtaattc gaatgacacc tatactgatg cagacaaaga gtcaccaca 1260
accgcattct acactatcaa ctacaacgct gataaggata ccggtacagt aactgtagct 1320
tcaaatggcg cagggtgcaac tggtaaatct gcagctactg ttggggcaca ggcttatgtt 1380
aactctacag gcaactgac cactgaaacc accagtgcag gcaactgcaac caaagatcct 1440
ctggctgccc tggatgaagc tatcagctcc atcgacaaat tccgttcac cctgggtgct 1500
atccagaacc gtctggatto cgcggttacc aacctgaaca acaccactac caacctgtcc 1560
gaagcgcagt cccgtattca ggacgcgac tatgcgaccg aagtgtccaa catgtcgaaa 1620
gcgcagatta tccagcaggc cggtaaactc gtgctggcaa aagccaacca ggtaccgcag 1680

```

cagggttctgt ctctgctgca ggggttaa

1707

<210> 44

<211> 1720

<212> DNA

<213> Escherichia coli

<400> 44

```
atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccggt ttacttctaa tattaaaggc 180
ctgactcagg ctgcacgtaa cgccaatgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtgtgctg aactgaccgt tcaggcgacc 300
accggtacca actcccagtc tgatctggac tctatccagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc cggtcagact cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcca tgaaaattca ggttggcgcg aatgatggcc agaccatcac tatcgacctg 480
aagaagattg actcttctac gttgaaactg actggtttta acgtgaatgg ttctggttct 540
gtggcgaata ctggggcgac taaagacgaa ctggctgctg ctgctgcggc ggcggttaca 600
actcctgctg tcgggtactga cggcgtgacc aaatataccg tagacgcagg gcttaacaaa 660
gccacagcag caaacgtgtt tgcaaacctt gcagatgggt ctggtgttga tgctagcatt 720
tccaacgggt ttggtgcagc agcagccaca gactacacct acaataaagc tacaatgat 780
ttactttca atgccagcat tgctgctggt gctgcggccg gtgatagtaa cagcgcagct 840
ctgcaatcct tcctgactcc aaaagcaggt gatacagcta acctgagcgt caaaatcggt 900
acgacatctg ttaatgttgt tctggcgagc gatggcaaaa ttacagcgaa agatgggtca 960
gctctgtata tcgactcaac gggtaacctg actcagaaca gcgcaggcac tgtaacagca 1020
gcaaccttgg atggactgac caaaaaccat gatgcgacag gagctggttg tgttgatata 1080
acgaccgcag atggcgcaac tatctctctg gcaggctctg ctaacgcggc aacagggtact 1140
caatcagggt caattacact gaaaaatgtt cgtatcagtg ctgatgctct gcagtctgct 1200
gcgaaaggta ctggtatcaa tgttgataat ggtgctgatg atatttctgt tagtaaaacc 1260
gggtgtcggt actaccggag gtgcgcctac ttatactgat gctgatggta aattaacgac 1320
aaccaacacc gttgattatt tcctgcaaac tgatggcagc gtaaccaatg gttctggtaa 1380
aggggtttac accgatgcag ctggtaaatt cactaccgac gctgcaacca aagccgcaac 1440
caccaccgat ccgctgaaag cccttgatga cgcaatcagc cagatcgata agttccgttc 1500
atccctgggt gctatccaga accgtctgga ttccgcgggt accaacctga acaacaccac 1560
taccaacctg tccgaagcgc agtcccgtat tcaggacgcc gactatgcga ccgaagtgtc 1620
caatatgtcg aaagcgcaga tcatccagca ggccggtaac tccgtgttgg caaaagctaa 1680
ccagggtacc cagcagggtt tgtctctgct gcagggttaa 1720
```

<210> 45

<211> 14516

<212> DNA

<213> Escherichia coli

<400> 45

```
gatctgatgg ccgtagggcg ctacgtgctt tctgctgata tctgggctga gttggaaaaa 60
actgctccag gtgcctgggg acgtattcaa ctgactgatg ctattgcaga gttggctaaa 120
aaacagtctg ttgatgcat gctgatgacc ggcgacagct acgactgcgg taagaagatg 180
ggctatatgc aggcattcgt taagtatggg ctgcgcaacc ttaaagaagg ggcgaagtgc 240
```

cgtaagagca tcaagaagct actgagtgag tagagattta cacgtctttg tgacgataag 300
 ccagaaaaaa tagcggcagt taacatccag gcttctatgc ttttaagcaat ggaatgttac 360
 tgccggttttt tatgaaaaat gaccaataat aacaagttaa cctaccaagt ttaatctgct 420
 ttttggttga ttttttcttg tttctggtcg catttggtta gacaattagc gtgagtttta 480
 gagagttttg cgggatctcg cggaactgct cacatctttg gcatttagtt agtgcactgg 540
 tagctgttaa gccaggggag gtagcttgcc taattaattt ttaacgtata catttattct 600
 tgccgcttat agcaataaaa gtcaatcgga ttaaacttct tttccattag gtaaaagagt 660
 gtttgtagtc gctcagggaa attgggtttg gtagtagtac ttttcaaatt atccattttc 720
 cgatttagat ggcagttgat gttactatgc tgcatacata tcaatgtata ttatttactt 780
 ttagaatgtg atatgaaaaa aatagtgatc ataggcaatg tagcgtcaat gatgttaagg 840
 ttcaggaaag aattaatcat gaatttagtg aggcaagggtg ataatgtata ttgtctagca 900
 aatgattttt ccactgaaga tcttaaagta ctttcgtcac ggggcgttaa ggggggttaa 960
 ttctctctta actcaaaggg tattaatcct ttttaaggata taattgctgt ttatgaacta 1020
 aaaaaaatc ttaaggatat ttcccagat attgtatttt catattttgt aaagccagta 1080
 atatttgga ctattgcttc aaagttgtca aaagtgccaa ggattgttgg aatgattgaa 1140
 ggtctaggta atgccttcac ttattataag ggaaagcaga ccacaaaaac taaaatgata 1200
 aagtggatc aaattctttt atataagtta gcattaccga tgcttgatga tttgattcta 1260
 ttaaatcatg atgataaaaa agatttaatc gatcagtata atattaaagc taaggtaaca 1320
 gtgttaggtg ggattggatt ggatcttaat gagttttcat ataaagagcc accgaaagag 1380
 aaaattacct ttatttttat agcaaggtta ttaagagaga aagggatatt tgagtttatt 1440
 gaagccgcaa agttcgtaa gacaacttat ccaagttctg aatttgtaat ttaggaggt 1500
 tttgagagta ataatccttt ctcatataaa aaaaatgaaa ttgaatcgct aagaaaagaa 1560
 catgatctta tttatcctgg tcatgtggaa aatgttcaag attggttaga gaaaagttct 1620
 gtttttggtt tacctacatc atatcgagaa ggcgtaccaa ggggtgatcca agaagctatg 1680
 gctattggta gacctgtaat aacaactaat gtacctgggt gtagggatat aataaatgat 1740
 ggggtcaatg gctttttgat acctccattt gaaattaatt tactggcaga aaaaatgaaa 1800
 tattttattg agaataaaga taaagtactc gaaatggggc ttgctggaag gaagtttgca 1860
 gaaaaaaaact ttgatgcttt tgaaaaaaat aatagactag catcaataat aaaatcaaat 1920
 aatgattttt gacttgagca gaaattattt atatttcaat ctgaaaaata aaggctgtta 1980
 ttatgaataa agtggcatta attactggta tcaactggga agatggctcc tatttggcag 2040
 aattattggt agaaaaaggt tatgaagttc atggtattaa acgccgtgca tcttcattta 2100
 atactgagcg agtggatcac atctatcagg attcacattt agctaatacct aaactttttc 2160
 tacactatgg cgatttgaca gatacttcca atctgacctg tattttaaaa gaagttcaac 2220
 cagatgaagt ttacaatttg ggggcgatga gccatgtagc ggtatcattt gagtccagcag 2280
 aatacactgc tgatgttgat gcgataggaa cattgcgtct tcttgaagct atcaggatat 2340
 tggggctgga aaaaaagaca aaattttatc aggcttcaac ttcagagctt tatgggttgg 2400
 ttcaagaaat tccacaaaaa gagactacgc cattttatcc acgttcgcct tatgctgttg 2460
 caaaattata tgcctatttg atcactgtta attatcgtga gtcttatggg atgtttgcct 2520
 gcaatggtat tctctttaac cacgaatcac ctgcgcgtgg cgagaccttt gttactcgta 2580
 aaataacacg cgggatagca aatattgctc aaggctctga taaatgctta tacttgggaa 2640
 atatggattc tctgcgtgat tggggacatg ctaaggatta tgtcaaaatg caatggatga 2700
 tgctgcagca agaaactcca gaagattttg taattgctac aggaattcaa tattctgtcc 2760
 gtgagtttgt cacaatggcg gcagagcaag taggcataga gtttagcattt gaaggtgagg 2820
 gagtaaatga aaaagggtgtt gttgtttcgg tcaatggcac tgatgctaaa gctgtaaacc 2880
 cgggcgatgt aattatatct gtagatccaa ggtattttag gcctgcagaa gttgaaacct 2940
 tgcttggcga tctactaat gcgcataaaa aattaggatg gagccctgaa attacattgc 3000
 gtgaaatggt aaaagaaatg gtttccagcg atttagcaat agcgaaaaag aacgtcttgc 3060
 tgaaagctaa taacattgcc actaatattc cgcaagaata aaaaagataa tacattaaat 3120

```

aattaaaaaat ggtgctagat ttatttagtac cattatTTTT ttttgggtga ctaatgttta 3180
ttacatcaga taaatttaga gaaattatca agttagttcc attagtatca attgatctgc 3240
taattgaaaa cgagaatggt gaatatTTTat ttggtcttag gaataatcga cgggccaaaa 3300
attatTTTTt tgttccaggt ggtaggattc gcaaaaatga atctattaaa aatgctTTTta 3360
aaagaatatc atctatggaa ttaggttaaag agtatggtat ttcaggaagt gTTTTaatg 3420
gtgtatggga acatttctat gatgatgggt ttttttctga aggcgaggca acacattata 3480
tagtgctttg ttacacactg aaagtTctta aaagtgaatt gaatctccca gatgatcaac 3540
atcgtgaata cctttggcta actaaacacc aaataaatgc taaacaagat gttcataact 3600
attcaaaaaa ttattTTTTt taattTTTTat taaaaattaa tatgcgagag aattgtatgt 3660
ctcaatgtct ttaccctgta attattgccg gaggaaccgg aagcctcta tggcctgtgt 3720
ctcgagtatt ataccctaaa caattTTTTaa atttagttgg ggattctaca atgttgcaaa 3780
caacaattac gcgtttggat ggcacTgaat gcgaaaatcc aattgttatc tgcaatgaag 3840
atcaccgatt tattgtagca gagcaattac gacagattgg taagctaacc aagaatatta 3900
tacttgagcc gaaaggccgt aatactgcac ctgccatagc tttagctgct tttatcgctc 3960
agaagaataa tcctaatagac gacctTTTat tattagtact tgcggcagac cactctataa 4020
ataatgaaaa agcattTcga gagtcaataa taaaagctat gccgtatgca acttctggga 4080
agttagtaac atttggaaatt attccggaca cggcaaatac tggttatgga tatattaaga 4140
gaagtTcttc agctgatcct aataaagaat tcccagcata taatgttgcg gagtttgtag 4200
aaaaaccaga tgttaaaaca gcacaggaat atattTcgag tgggaattat tactggaata 4260
gcggaatgtt tttattTcgc gccagtaaatt atcttgatga actacggaaa ttagaccag 4320
atatttatca tagctgtgaa tgtgcaaccg ctacagcaaa tatagatatg gactttgtcc 4380
gaattaacga ggctgagttt attaatgtgc ctgaagagtc tatcgattat gctgtgatgg 4440
aaaaaacaaa agacgtgtga gttctTccga tagatattgg ctggaatgac gtgggttctt 4500
ggTcatcact ttgggatata agccaaaagg attgccatgg taatgtgtgc catggggatg 4560
tgctcaatca tgatggagaa aatagtTTTta tttactctga gtcaagtctg gttgcgacag 4620
tcggagtaag taatttagta attgtccaaa ccaaggatgc tgtactggtt gcggaccgtg 4680
ataaagtcca aaatgtTaaa aacatagtTg acgatctaaa aaagagaaaa cgtgctgaat 4740
actacatgca tcgtgcagtt tttcgccctt ggggtaaatt cgatgcaata gaccaaggcg 4800
atagatatag agtaaaaaaa ataatagtta aaccaggaga agggTtagat ttaaggatgc 4860
atcatcatag ggcagagcat tggattgttg tatccggtac tgctaaagtt tcaactaggta 4920
gtgaagttaa actattagtT tctaattgagT ctatatatat ccctcaggga gcaaaatata 4980
gtcttgagaa tccaggcgta atacctttgc atctaattga agtaagtTct ggtgattacc 5040
ttgaatcaga tgatatagtT cgTTTTactg acagatataa cagtaaaca ttcctaaagc 5100
gagattgata aatatgaata aaataactTg cttcaaagca tatgatatac gtgggcgtct 5160
tggtgctgaa ttgaatgatg aaatagcata tagaattggt cgcgcttatg gtgagttttt 5220
taaacctcaa actgtagtTg tgggaggaga tgctcgctta acaagtgaga gtttaaagaa 5280
atcactctca aatgggctat gtgatgcagg cgtaaagtgc ttagatcttg gaatgtgtgg 5340
tactgaagag atatatTTTT ccactTggta tttaggaatt gatggTggaa tcgaggtaac 5400
tgcaagccat aatccaattg attataatgg aatgaaatta gtaaccaaag gtgctcgacc 5460
aatcagcagt gacacaggTc tcaaagatat acaacaatta gtagagagta ataattttga 5520
agagctcaac ctagaaaaaa aagggaatat taccaaatat tccacccgag atgcctacat 5580
aatcattTg atgggctatg ctaatctgca aaaaataaaa aaaatcaaaa tagttgtgaa 5640
ttctgggaat ggtgcagctg gtctgttat tgatgctatt gaggaatgct ttttacggaa 5700
caatattccg attcagttTg taaaaataaa taatacacc gatggtaatt tccacatgg 5760
tatccctaatt ccattactac ctgagtgcag agaagatacc agcagtgcgg ttataagaca 5820
tagtgctgat tttggtattg catttgatgg tgattttgat aggtgttttt tctttgatga 5880
aaatggacaa tttattgaag gatactacat tgttggTTTta ttagcggaag tttttttagg 5940
gaaatatcca aacgcaaaaa tcattcatga tctcgcctt atatggaata ctattgatat 6000

```

cgtagaaaagt catggtggta tacctataat gactaaaacc ggtcatgctt acattaagca 6060
 aagaatgcgt gaagaggatg ccgtatatgg cggcgaaatg agtgcgcatc attattttta 6120
 agatttttgca tactgcgata gtggaatgat tccttggatt ttaatttgtg aacttttgag 6180
 tctgacaaat aaaaaattag gtgaactggg ttgtgggtgt ataaacgact ggccggcaag 6240
 tggagaaaata aactgtacac tagacaatcc gcaaaatgaa atagataaat tatttaaatcg 6300
 ttacaaagat agtgccttag ctgttgatta cactgatgga ttaactatgg agttctctga 6360
 ttggcggtttt aatgttagat gctcaaatac agaacctgta gtacgattga atgtagaatc 6420
 taggaataat gctattctta tgcaggaaaa aacagaagaa attctgaatt ttatatcaaa 6480
 ataaatttgc acctgagttc ataatgggaa caagaaatat atgaaagtac ttctgactgg 6540
 ctcaactggc atggttggta agaatatatt agagcatgat agtgcaagta aatataatat 6600
 acttactcca accagctctg atttgaattt attagataaa aatgaaatag aaaaattcat 6660
 gcttatcaac atgccagact gtattataca tgcagcggga ttagttggag gcattcatgc 6720
 aaatataagc aggccgtttg attttctgga aaaaaatttg cagatgggtt taaatttagt 6780
 ttccgtcgca aaaaaactag gtatcaagaa agtgcttaac ttgggtagtt catgcatgta 6840
 ccccaaaaac tttgaagagg ctattcctga gaaagctctg ttaactgggt agctagaaga 6900
 aactaatgag ggatatgcta ttgcgaaaat tgctgtagca aaagcatgcg aatataatc 6960
 aagagaaaac tctaattatt ttataaaaac aattatccca tgtaatttat atgggaaata 7020
 tgataaattt gatgataact cgtcacatat gattccggca gttataaaaa aaatccatca 7080
 tgcgaaaatt aataatgtcc cagagatcga aatttggggg gatggtaatt cgcgccgtga 7140
 gtttatgtat gcagaagatt tagctgatct tattttttat gttattccta aaatagaatt 7200
 catgccta atggttaa atgctgttagg ttacgattat tcaattaatg actattataa 7260
 gataattgca gaagaaattg gttatactgg gagtttttct catgatttaa caaaaccaac 7320
 aggaatgaaa cggaagctag tagatatctt attgcttaat aaaattgggt ggtcaagtca 7380
 ctttgaactc agagatggca tcagaaagac ctataattat tacttggaga atcaaaataa 7440
 atgattacat acccacttgc tagtaatact tgggatgaat atgagtatgc agcaatacag 7500
 tcagtaattg actcaaaaat gtttaccatg ggtaaaaagg ttgagttata tgagaaaaat 7560
 tttgctgatt tgtttggtag caaatatgcc gtaatgggtta gctctgggtc tacagctaat 7620
 ctgttaatga ttgctgccct tttcttcact aataaaccaa aacttaaaag aggtgatgaa 7680
 ataatagtac ctgcagtgtc atggtctacg acatattacc ctctgcaaca gtatggctta 7740
 aaggtgaagt ttgtcgatat caataaagaa actttaaata ttgatatcga tagtttgaaa 7800
 aatgctattt cagataaaaac aaaagcaata ttgacagtaa atttattagg taatccta at 7860
 gattttgcaa aaataaatga gataataaat aatagggata ttatcttact agaagataac 7920
 tgtgagtcga tgggcgcggt ctttcaaaaat aagcaggcag gcacattcgg agttatgggt 7980
 acctttagtt ctttttactc tcatcatata gctacaatgg aagggggctg cgtagttact 8040
 gatgatgaag agctgtatca tgtattgttg tgccttcgag ctcatgggtg gacaagaaat 8100
 ttacaaaag agaatatggt tacaggcact aagagtgatg atattttcga agagtcgttt 8160
 aagtttggtt taccaggata caatgttcgc ccacttgaaa tgagtgggtg tattgggata 8220
 gagcaactta aaaagttacc aggttttata tccaccagac gttccaatgc acaatatatt 8280
 gtagataaat ttaaagatca tccattcctt gatatacaaa aagaagttgg tgaaagtagc 8340
 tggtttggtt tttccttcgt tataaaggag ggagctgcta ttgagaggaa gagtttagta 8400
 aataatctga tctcagcagg cattgaatgc cgaccaattg ttactgggaa ttttctcaaa 8460
 aatgaacgtg ttttgagtta ttttgattac tctgtacatg atacggtagc aaatgccgaa 8520
 tatatagata agaattggtt ttttgcgga aaccaccaga tacctttggt taatgaaata 8580
 gattatctac gaaaagtatt aaaataacta acgaggcact ctatttcgaa tagagtgcct 8640
 ttaagatggg attaacagtg aaaaaaattt tagcgtttgg ctattctaaa gtactaccac 8700
 cggttattga acagtttgtc aatccaattt gcatcttcat tatcacacca ctaataactca 8760
 accacctggg taagcaaagc tatggtaatt ggattttatt aattactatt gtatcttttt 8820
 ctcagttaat atgtggagga tgttcgcgat ggattgcaaa aatcattgca gaacagagaa 8880

| | | | | | | |
|-------------|-------------|------------|------------|-------------|-------------|-------|
| ttcttagtga | tttatcaaaa | aaaaatgctt | tacgtcaaat | ttcctataat | ttttcaattg | 8940 |
| ttattatcgc | atttgccgta | ttgatttctt | ttcttatatt | aagtatttgt | ttcttcgatg | 9000 |
| ttgcgaggaa | taattcttca | ttcttattcg | cgattattat | ttgtgggttt | tttcaggaag | 9060 |
| ttgataat | atttagtggt | gcgctaaaag | gttttgaaaa | atttaaatgta | tcattgtttt | 9120 |
| ttgaagtaat | tacaagagtg | ctctgggctt | ctatagtaat | atatggcatt | tacggaaatg | 9180 |
| cactcttata | ttttacatgt | ttagccttta | ccattaaagg | tatgctaaaa | tatattcctg | 9240 |
| tatgtctgaa | tattaccggt | tgtttcatca | atcctaattt | taatagagtt | gggattgtta | 9300 |
| atttgttaaa | tgagtcaaaa | tggtatgttc | ttcaattaac | tggtggcgtc | tcacttagtt | 9360 |
| tgtttgatag | gctcgtaata | ccattgattt | tatctgtcag | taaactggct | tcttatgtcc | 9420 |
| cttgccctca | actagctcaa | ttgatgttca | ctctttctgc | gtctgcaa | caaataattac | 9480 |
| taccaatggt | tgctagaatg | aaagcatcta | acacatttcc | ctctaattgt | ttttttaaaa | 9540 |
| ttctgcttgt | atcactaatt | tctgttttgc | cttgtcttgc | gttattcctt | tttggtcgtg | 9600 |
| atatattatc | aatatggata | aaccctacat | ttgcaactga | aaattataaa | ttaatgcaaa | 9660 |
| ttttagctat | aagttacatt | ttattgtcaa | tgatgacatc | ttttcatttc | ttgttattag | 9720 |
| gaattggtaa | atctaagctt | gttgcaaatt | taaatctggt | tgcagggctc | gcacttgctg | 9780 |
| cttcaacggt | aatcgagct | cattatggcc | tttatgcaat | atctatggta | aaaataatat | 9840 |
| atccggcttt | tcaattttat | tacctttatg | tagcttttgt | ctattttaat | agagcgaaaa | 9900 |
| atgtctattg | atttactttt | ttcaattact | gaaatcgcaa | ttgttttttc | ttgcactatt | 9960 |
| tacatattta | ctcaatgttt | gttaatgcgg | aggatctatt | tagataaaaag | tatttttaatt | 10020 |
| cttttatgct | tgctcttttt | tttagtaatc | attcaacttc | ctgagcttaa | tgtaaacggt | 10080 |
| ttggctgatt | ctttaaagtt | atcactgcct | ttattgatgg | tctttatcgc | ttttcaaaaa | 10140 |
| ccgaaattat | gcttgtgggt | tattattgca | ttgttgtttt | tgaactctgc | atttaatttt | 10200 |
| ttatatttaa | agacattcga | taagtttagc | tcatttcctt | ttactttttt | tatattgctg | 10260 |
| ttttacttgt | ttagattggg | aattggtaat | ttaccggttt | ataaaaaata | aaaattttac | 10320 |
| gcgttgattt | ttctctttat | attaatagac | ataatgcagt | cattgttaat | aaattatagg | 10380 |
| gggcagattt | tatatccggt | aatttgcac | ctgatacttg | tgtttaaaagt | taatttaaga | 10440 |
| aaaaagattc | catacttttt | tttaatgctg | ccagttttat | atgtaattat | tatggcttat | 10500 |
| attgggtttta | attattttcaa | taaaggcgta | actttttttg | aacctacagc | aagtaatatt | 10560 |
| gaacgtacgg | ggatgatata | ttatttggtt | tcacagcttg | gtgattatat | attccatggt | 10620 |
| atggggacat | taaattttct | aaataacggc | ggacaatata | agacgttata | tggacttcca | 10680 |
| tcattaattc | ctaatagacc | tcattgattt | ttattacggt | tctttataag | tattggtgtg | 10740 |
| ataggagcat | tggtttatca | ttctatat | tttgtttttt | ttaggagaat | atctttctta | 10800 |
| ttatatgaga | gaaatgctcc | tttcattggt | gtaagtgtgt | tgttactgtt | acaagtgtgt | 10860 |
| tttaatttata | cattaaaccc | ttttgatgct | tttaatcgat | tgatttgcgg | gcttacagtt | 10920 |
| ggagttgttt | atggatttgc | aaaaattaga | taagtatacc | tgtaatggaa | atttagacgc | 10980 |
| tccacttggt | tcaataatca | ttgcaactta | taattctgaa | cttgatatag | ctaagtgttt | 11040 |
| gcaatcggt | actaatcaat | cttataagaa | tattgaaatc | ataataatgg | atggaggatc | 11100 |
| ttctgataaa | acgcttgata | ttgcaaaatc | gtttaaagac | gaccgaataa | aaatagtttc | 11160 |
| agagaaagat | cgtggaattt | atgatgcctg | gaataaagca | gttgatttat | ccattggtga | 11220 |
| ttgggtagca | tttattgggt | cagatgatgt | ttactatcat | acagatgcaa | ttgcttcatt | 11280 |
| gatgaagggg | gttatgggtat | ctaattggcg | ccctgtgggt | tatgggagga | cagcgcacga | 11340 |
| aggtcccgat | aggaacatat | ctggattttc | aggcagtga | tggtacaacc | taacaggatt | 11400 |
| taagtttaat | tattacaaat | gtaatttacc | attgcccatt | atgagcgcaa | tatattctcg | 11460 |
| tgatttcttc | agaaacgaac | gttttgatat | taaattaaaa | attgttgctg | acgctgattg | 11520 |
| gtttctgaga | tgtttcatca | aatggagtaa | agagaagtca | ccttatttta | ttaatgacac | 11580 |
| gacccttatt | gttagaatgg | gatatggtgg | ggtttcgact | gatatttctt | ctcaagttaa | 11640 |
| aactacgcta | gaaagtttca | ttgtacgcaa | aaagaataat | atatactgtt | taaacataca | 11700 |
| gctgattctt | agatatgcta | aaattctggt | gatggtagcg | atcaaaaaata | tttttggtgaa | 11760 |

taatgtttat aaattaatgc ataacgggta tcattcccta aagaaaatca agaataaaat 11820
atgaagattg tttatataat aaccgggctt acttgtggtg gagccgaaca cttatgacg 11880
cagtttagcag accaaatggt tatacgcggt catgatgtta atattatttg tctaactggt 11940
atatctgagg taaagccaac acaaaatatt aatattcatt atgttaatat ggataaaaaat 12000
tttagaagct tttttagagc tttttttcaa gtaaaaaaaaa taattgtcgc cttaaagcca 12060
gatataatac atagtcatat gtttcatgct aatattttta gtcgttttat taggatgctg 12120
attccagcgg tgccctgat atgtaccgca cacaacaaaa atgaagggtg caatgcaagg 12180
atgttttgtt atcgactgag tgatttttta gcttctatta ctacaaatgt aagtaaagag 12240
gctgttcaag agtttatagc aagaaaggct acacctaaaa ataaaatagt agagattccg 12300
aattttatta atacaaataa atttgatttt gatattaatg tcagaaagaa aacgcgagat 12360
gcttttaatt tgaaagacag tacagcagta ctgctcgcag taggaagact tggtgaagca 12420
aaagactatc cgaacttatt aaatgcaata aatcatttga ttctttcaaa aacatcaaat 12480
tgtaatgatt ttattttgct tattgctggc gatggcgcat taagaaataa attattggat 12540
ttggtttgtc aattgaatct tgtggataaa gttttcttct tggggcaaag aagtgatatt 12600
aaagaattaa tgtgtgctgc agatcttttt gttttgagtt ctgagtggga aggttttggg 12660
ctcgttgttg cagaagctat ggcggtgtgaa cgtcccggtg ttgctaccga ttctggtgga 12720
gttaaagaag tcgttgacc tcataatgat gttatccctg tcagtaatca tattctgttg 12780
gcagagaaaa tcgctgagac acttaaaata gatgataacg caagaaaaat aataggtatg 12840
aaaaatagag aatatattgt ttccaatttt tcaattaaaa cgatagttag tgagtgggag 12900
cgcttatatt ttaaattatt caagcgtaat aatataattg attgaaata taagtttgta 12960
ctctggatgc aatagtttct ctatgctgtt tttttactgg ctccgtattt ttacttatag 13020
ctggattttg ttatatatca gtattaatct gtctcaactt catctagact acattcaagc 13080
cgcgcatgcg tcgcgcggtg actacacctg acaggagtat gtaatgtcca agcaacagat 13140
cggcgctcgc ggtatggcag tgatggggcg caacctggcg ctcaacatcg aaagccgcgg 13200
ttataccgct tccatcttca accgctcccg cgagaaaact gaagaagttg ttgccgagaa 13260
cccggataag aaactgggtc cttattacac ggtgaaagag ttctgctgagt ctcttgaaac 13320
cccacgctcg atcctgttaa tggtaaaagc agggggcgga actgatgctg ctatcgattc 13380
cctgaagccg tatctggata aaggcgacat cattattgat ggtggcaaca cttcttcca 13440
ggacactatc cgtcgtaacc gtgaactgtc cgcggaaggc tttaacttca tcggtaccgg 13500
cgtgtccggc ggtgaagagg gcgcctgaa aggcccatct atcatgccag gtggccagaa 13560
agaagcgat gagctggtg cgccatcct gaccaagatt gctgcggtt ctgaagatgg 13620
cgaaccatgt ataacttaca tcggtgctga cgggtgcgggt cactacgtga agatggtgca 13680
caacggatc gaatatggcg atatgcagct gattgctgaa gcctattctc tgcttaaagg 13740
cggccttaat ctgtctaacg aagagctggc aaccactttt accgagtgga atgaaggcga 13800
gctaagtagc tacctgattg acatcaccaa agacatcttc accaaaaaag atgaagaggg 13860
taaataacct gttgatgtga tcctggacga agctgcgaac aaaggcaccg gtaaatggac 13920
cagccagagc tctctggatc tgggtgaacc gctgtcgtg atcacgaat ccgtattcgc 13980
tcgctacatc tcttctctga aagaccagcg cattgcggca tctaaagtgc tgtctggtcc 14040
gcaggctaaa ctggctggtg ataaagcaga gttcgttgag aaagtccgtc gcgcgctgta 14100
cctgggtaaa atcgtctctt atgcccaagg cttctctcaa ctgctgcccg cgtctgacga 14160
atacaactgg gatctgaact acggcgaaat cgcggaagatc ttccgcgcgg gctgcatcat 14220
tcgtgcgcag ttctgcaga aaattactga cgcgatgct gaaaacaaag gcattgctaa 14280
cctgttgctg gtcctgtact tcaaaaatat cgctgatgaa tatcagcaag cgctgcgtga 14340
tgtagtggct tatgctgtgc agaacgggat tccggtaccg accttctctg cagcggtagc 14400
ctactacgac agctaccgtt ctgcggtact gccggcta atctgattcagg cacagcgtga 14460
ttacttcggt gcgcacacgt ataaacgcac tgataaagaa ggtgtgttcc acaccg 14516

<210> 46

<211> 1380

<212> DNA

<213> Escherichia coli

<400> 46

```
aacaaatctc agtcttctct tagctctgct attgagcgtc tgtcttctgg tctgcgtatt 60
aacagcgcaa aagacgatgc agcaggctcag gcgattgcta accgttttac ggcaaattatt 120
aaagggtctga cccaggcttc ccgtaacgcg aatgatggta tttctgttgc gcagaccact 180
gaagggtgcgc tgaatgaaat taacaacaac ctgcagcgta ttcgtgaact ttctgttcag 240
gcaactaacg gtactaactc tgacagcgat ctttcttcta tccaggctga aattactcaa 300
cgtctggaag aaattgaccg tgtatctgag caaactcagt ttaacggcgt gaaagtcctt 360
gctgaaaata atgaaatgaa aattcagggtt ggtgctaata atggtgaaac catcactatc 420
aatctggcaa aaattgatgc gaaaactctc ggctggacg gttttaatat cgatggcgcg 480
cagaaagcaa cgggcagtga cctgatttct aaatttaaag cgacaggtaac tgataattat 540
caaattaacg gtactgataa ctatactgtt aatgtagata gtggagtagt acaggataaa 600
gatggcaaac aagtttatgt gaggctgcg gatggttcac ttacgaccag cagtgatact 660
caattcaaga ttgatgcaac taagcttgca gtggctgcta aagatttagc tcaaggtaat 720
aagattgtct acgaagggtat cgaatttaca aataccggca ctggcgctat acctgccaca 780
ggtaatggtg aattaaccgc caatgttgat ggtaaggctg ttgaattcac tatttcgggg 840
agtgtgata catcagggtac tagtgcaacc gttgcccta cgacagccct atacaaaaat 900
agtgcagggc aattgactgc aacaaaagtt gaaaataaag cagcgacact atctgatctt 960
gatctgaacg ctgccaagaa aacaggaagc acgttagttg ttaacgggtgc aacttacgat 1020
gttagtgagc atggtaaaac gataacggag actgcttctg gtaacaataa agtcatgtat 1080
ctgagcaaat cagaagggtg tagcccgatt ctggtaaacg aagatgcagc aaaatcgttg 1140
caatctacca ccaaccgct cgaaactatc gacaaagcat tggctaaagt tgacaatctg 1200
cgttctgacc tcggtgcagt acaaaaccgt ttcgactctg ccatcaccaa ccttggaac 1260
accgtaaaca acctgtcttc tgcccgtagc cgtatcgaag atgctgacta cgcgaccgaa 1320
gtgtctaaca tgtctcgtgc gcagatcctg caacaagcgg gtacctctgt tctggcacag 1380
```

<210> 47

<211> 1497

<212> DNA

<213> Escherichia coli

<400> 47

```
atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gtaaccgtt tcacctctaa cattaaggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtatctccg ttgcgcagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtgtgcgtg aactgacggg acaggccact 300
accggtacta actctgagtc tgatctgtct tctatccagg acgaaattaa atcccgtctg 360
gatgaaattg accgcgtatc tggtcagacc cagttcaacg gcgtgaacgt gctggcaaaa 420
aatggctcca tgaaaatcca ggttggcgca aatgataacc agactatcac tatcgatctg 480
aagcagattg atgctaaaac tcttggcctt gatggtttta gcgttaaaaa taacgatata 540
gttaccacta gtgctccagt aactgctttt ggtgctacca ccacaaacaa tattaaactt 600
actggaatta cctttctac ggaagcagcc actgatactg gcggaactaa ccagcttca 660
attgaggggtg tttatactga taatggtaat gattactatg cgaaaatcac cggtggtgat 720
aacgatggga agtattacgc agtaacagtt gctaatagat gtacagtgc aatggcgact 780
```

ggagcaacgg caaatgcaac tgtaactgat gcaaatacta ctaaagctac aactatcact 840
tcaggcggta cacctgttca gattgataat actgcagggt cgcgaactgc caaccttgg 900
gctgttagct tagtaaaact gcaggattcc aagggtaatg ataccgatac atatgcgctt 960
aaagatacaa atggcaatct ttacgctgcg gatgtgaatg aaactactgg tgctgtttct 1020
gttaaaacta ttacctatac tgactcttcc ggtgccgcca gttctccaac cgcgggtcaaa 1080
ctgggaggag atgatggcaa aacagaagtg gtcgatattg atggtaaaac atacgattct 1140
gccgatttaa atggcggtaa tctgcaaaca ggtttgactg ctgggtggtga ggctctgact 1200
gctgttgcaa atggtaaaac cacggatccg ctgaaagcgc tggacgatgc tatcgcatct 1260
gtagacaaat tccgttcttc cctcgggtgcg gtgcaaaacc gtctggattc cgcggttacc 1320
aacctgaaca acaccactac caacctgtct gaagcgcagt cccgtattca ggacgccgac 1380
tatgcgaccg aagtgtccaa tatgtcgaaa gcgcagatca tccagcaggc cggtaactcc 1440
gtgttggtgcaa aagctaacca ggtaccgcag caggttctgt ctctgctgca ggggttaa 1497

<210> 48

<211> 1695

<212> DNA

<213> Escherichia coli

<400> 48

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
gcgctgtctg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctg 360
gacgaaattg accgcgtatc cgggtcaaacc cagttcaacg gtgtgaacgt actggcgaaa 420
gacgggttcga tgaaaattca ggttggtgcg aatgacggcc agactatcac tattgatctg 480
aagaaaattg actctgatac gctggggctg aatgggttta acgttaacgg caaagggtact 540
attgcgaaca aagcggcaac cattagtgat ctggcggcga cgggggcgaa tgttactaac 600
tcaagcaata ttgttgtcac gacaaagttc aatgccttgg atgcagcgac tgcatttagc 660
aaactcaaag atggtgattc tgttgccgtt gctgctcaga aatatactta taacgcatcg 720
accaatgatt ttacgacaga aaatacagta gcgacaggca ctgcaacgac agatcttggc 780
gctactctga aggtgtctgc tgggcagagt caatcaggta catatacctt tgcaaatggt 840
aaagtttaact ttgatgttga tgcaagcggg aatatcacta ttggcggcga aaaggcttct 900
ttgggttggtg gagcgctgac tactaacgat cccaccggct ccactccagc aacgatgtct 960
tcctgttcta aggcgcgga tgacaaagat gccgtcaat cctcgattga ttttggcggg 1020
aaaaaatacg aatttgctgg tggcaattct actaatggtg gcggcggtta attcaaagac 1080
acgggtgtctt ctgacgcgct tttggctcag gttaaagcgg atagtactgc taataatgta 1140
aaaatcacct ttaacaatgg tcctctgtca ttactgcat cgttccaaa tgggtgtatct 1200
ggctccgcgg catcgatgc agcctacatt gatagcgaag gcgaactgac aactactgaa 1260
tcctacaaca caaattattc cgtagacaaa gacacggggg ctgtaagtgt tacagggggg 1320
agcggtagcg gtaaatacgc cgcaaactg ggtgctcagg cttatgtagg tgcagatggt 1380
aaattaacca cgaatactac tagtaccggc tctgcaacca aagatccact aaatgcgctg 1440
gatgaggcaa ttgcatccat cgacaaattc cgttcttccc tgggggctat ccagaaccgt 1500
ctggattccg cagtcaccaa cctgaacaac accactacca acctgtctga agcgcagtc 1560
cgtattcagg acgccgacta tgcgaccgaa gtgtccaaca tgtcgaaagc gcagatcatc 1620
cagcaggccg gtaactccgt gttggcaaaa gctaaccagg taccgcagca ggttctgtct 1680
ctgctgcagg gtttaa 1695

- 37 -

<210> 49

<211> 1164

<212> DNA

<213> Escherichia coli

<400> 49

```

aacaagaacc agtctgcgct gtcgagttct atcgagcgct tgtcttctgg cttgcggtatt 60
aacagcgcgga aggatgacgc cgcgggtcag gcgattgcta accgttttac ttctaacatt 120
aaaggcctga ctcaggctgc acgtaacgcc aacgacggta tttctgttgc gcagaccacc 180
gaaggcgcgc tgtccgaaat taacaacaac ttacagcgctg tgcgtgagct gactgttcag 240
gcgaccaccg gtactaactc tgagtctgac ctgtcttcta tccaggacga aatcaaactc 300
cgcctggaag agattgatcg tgtttcaagt cagactcaat ttaacggcgt gaatgttttg 360
gctaaagatg ggaaaatgaa cattcagggtt ggggcaagtg atggacagac tatcactatt 420
gatctgaaaa agatcgattc atctacacta aacctctcca gttttgatgc taaaaacttg 480
ggcaccagtg ttaaagatgg ggccaccatc aataagcaag tggcagtaga tgctggcgac 540
tttaaagata aagcttcagg atcgttagg accctaaaat tagttgagaa agacggtaag 600
tactatgtaa atgacactaa aagtagtaag tactacgatg ccgaagtaga tactagtaag 660
ggtgaaaatta acttcaactc taaaaatgaa agtggaacta ctctactgc agcgacggaa 720
gtaactactg ttggccgcgga tgtaaaattg gatgcttctg cacttaaagc caaccaatcg 780
cttgtcgtgt ataaagataa aagcggcaat gatgcttata tcattcagac caaagatgta 840
acaactaatc aatcaacttt caatgccgct aatatcagtg atgctgggtg tttatctatt 900
ggtgcatcta caaccgcgcc aagcaattta acagctgacc cgcttaaggc tcttgatgat 960
gcaattgcat ctgttgataa attccgctct tctctcgggt ccgttcagaa ccgtctggat 1020
tctgccattg ccaacctgaa caacaccact accaacctgt ctgaagcgca gtcccgtatt 1080
caggacgctg actatgcgac cgaagtgtcc aacatgtcga aagcgcagat tatccagcag 1140
gccggttaact ccgtgctggc aaaa

```

<210> 50

<211> 1818

<212> DNA

<213> Escherichia coli

<400> 50

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccgtt tcacctctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgctaacgat ggtatctctc tggcgcagac cactgaaggc 240
gcactgtctg agattaacaa caacttaciaa cgtgtgctg agttgactgt acaggcgacc 300
accggtacta actctgattc tgacctggct tctattcagg acgaaatcaa atcccgtttg 360
tctgaaattg accgcgtatc cgggcagacc cagttcaacg gcgtgaacgt attgtctaaa 420
gatggctccc tgaaaattca ggttggecga aatgatggtc agactatctc tatcgacctg 480
aagaaaattg actctgatac tctgggtttg aatgggttca acgttaatgg ttctggtacc 540
attgcaaaca aagcggccac aatcagtgac ttgactgctc agaaagccgt tgacaacggc 600
aatggtactt ataaagttac aactagcaac gctgcactta ctgcatctca ggcattaagt 660
aagctgagtg atggcgatac tgtagatatt gcaacctatg ctggtggtac aagttcaaca 720
gttagttata aatacgacgc agatgcaggt aacttcagtt ataacaatac tgcaaacaaa 780
acaagtgtg cggctggaac tctggcagat actcttctcc cggcagctgg ccagactaaa 840

```

accggtactt acaaggctgc tactggtgat gttaacttta atgttgacgc aactggtaat 900
ctgacaattg gcggacagca agcctacctg actactgatg gtaaccttac aacaaacaac 960
tccggtgggtg cggctactgc aactcttaaa gagctgttta ctcttgctgg cgatggtaaa 1020
tctctgggga acggcggtac tgctaccgtt actctggata atactacgta taatttcaaa 1080
gctgctgcga acgttactga tgggtgctggt gtcacgctg ctgctggtgt aacttatata 1140
gccactgttt cttaaagatgt cattctggca caactgcaat ctgcaagtca ggcagcagca 1200
accgctaccg acggtgatac tgtcgcaacg atcaactata aatctggtgt catgatcggg 1260
tccgctacct ttaccaatgg taaagggtact gccgatggta tgacttctgg tacaactcca 1320
gtcgtagcta cagggtgctaa agctgtatat gttgatggca acaatgaact gacttccact 1380
gcatcttacg atacgactta ctctgtcaac gcagatacag gcgcagtaaa agtggatatca 1440
ggtactggta ctggtaaat tgaagctggt gctggtgcgg atgcttatgt aagcaaagat 1500
ggcaaattaa cgacagaaac caccagtgcg ggcactgcaa ccaaagatcc tttggctgcc 1560
ctggatgctg ctatcagctc catcgacaaa ttccgttccct ccctgggtgc tatccagaac 1620
cgtctggatt ccgcagtcac caacctgaac aacaccacta ctaacctgtc tgaagcgag 1680
tcccgatttc aggacgccga ctatgcgacc gaagtgtcca atatgtcgaa agcgagatc 1740
atccagcagg ccggtaaactc tgtgttggca aaagctaacc aggtaccgca gcaggttctg 1800
tctctgctgc aggggttaa 1818

<210> 51

<211> 1344

<212> DNA

<213> Escherichia coli

<400> 51

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggg tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
gacgaaattg accgcgtttc cggtcagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacggttcga tgaagattca ggttggcgcg aatgacgggc agaccatctc tatcgatttg 480
cagaaaattg attcttcaac gctgggattg aaagggttct cggtatcagg gaacgcatta 540
aaagttagcg atgcgataac tacagttcct ggtgctaatt ctggcgatgc cccggttacg 600
gttaaatttg gtgcgaacga taccgctgct gccgcaatgg ctaaaacatt gggaataagt 660
gatacatcag gcttgtccct acataacgta caaagcgcg atggtaaagc gacaggaacc 720
tatgttggtc aatctggtaa tgacttctat tcggcttccg ttaatgctgg tggcgttggt 780
acgcttaata ccaccaatgt tactttcact gatcctgcga acggtgttac cacagcaaca 840
cagacaggtc agcctatcaa ggtcacgacg aatagtgtct gcgcggctgt tggctatggt 900
actattcaag gcaaagatta ccttgcgtgt gcagacggta aggatgcaat tgaaaacggg 960
ggtgacgctg caacaaatga agacacaaaa atccaactta ccgatgaact cgatgttgat 1020
ggttctgtaa aaacagcggc aacagcaaca ttttctggta ctgcaaccaa cgatccgctg 1080
gcacttttag acaaagctat ctgcgaagtt gatactttcc gtcctccct cggtgccgta 1140
caaaaccgtc tggattctgc ggtcaccaac ctgaataaca ccaccacaa cctgtctgaa 1200
gcgcagtccc gtattcagga cgcgcactat gcgaccgaag tgtccaacat gtcgaaagcg 1260
cagatcatcc agcaggcggg taactctgtg ctgtctaaag ctaaccaggg accgcagcag 1320
gttctgtctc tgctgcaggg ttaa 1344

<210> 52

<211> 2599

<212> DNA

<213> Escherichia coli

<400> 52

cttctcttag ctctgctatt gagcgtctgt cttctggtct gcgtattaac agcgcaaaag 60
acgatgcagc aggtcaggcg attgctaacc gttttacggc aaatattaaa ggtctgaccc 120
aggcttcccg taacgcgaat gatggtatct ctggtgcgca gaccactgaa ggtgcgctga 180
atgaaattaa caacaacctg cagcgtattc gtgaactttc tgttcaggca actaacggta 240
ctaactctga cagcgtatct tcttctatcc aggctgaaat tactcaacgt ctggaagaaa 300
ttgaccgtgt atctgagcaa actcagttta acggcgtgaa agtccttgct gaaaataatg 360
aaatgaaaat tcaggttggt gctaattgat gtgaaacat tgacctgccc ccacgattag 420
atacaacact cagttagtaa cgtcgggaatc ttcattctca gaatgacct tctccagcc 480
cgctgcaaat tcagacgggt tctgataatt cagcgtggag tgcggggcggc attcggtata 540
atcctgcgcg cagtcattaa taattttcct ggcatgaacg atatcgctga accagtgtc 600
attcaaacat tcacgcgcaa atcgtccgtt aaagctctca ataaatccgt tctgcgttg 660
cttgcccggc tggattaagc gcaactcaac accatgctca aaggccatt gatccagtgc 720
acggcaagtg aactccggcc cctggtcagt tcttatcgct gccggatagc ctcgaaacag 780
tgcaatgctg tccagaatac gcgtgacctg aacgcctgaa atcccaaagg caacagtgc 840
cgtcaggcat tctttgtga aatcatcgac gcaggtaaga cacttgatcc tgcgaccggt 900
ggaaagtgcg tccatgacga aatccatcga ccaggtcaga ttggggcgccg ccggacggag 960
cagcggcaga cgttctgttg ccagcccttt acgacgtctt ctgcgtttta cgcccaggcc 1020
actgaggtga taaagccggc acacgcgctt atgattaaca tgaagccctt cagggcgag 1080
caactgccaa atacgacggt agccaaaacg cctgcgctcc agtgccagct cagtgtgctg 1140
ccctgataaa tgcgcatcag cagccggacg gtgagcctca tagcggcagg tcgacaggga 1200
taaacctgta agcctgcagg cacgacgttg cgacagaccg gtcgcatcac acatcaacat 1260
cacggcttcc cgcttctggt ctgtcgtcag tactttcgcc caagagccac ctgaagcgcc 1320
tctttatcca gcatggcttc ggcaagcagc ttcttgagtc tgggtgttctc ttcctcaagc 1380
gacttcaggc gcttaacttc aggcacctcc ataccgccat acttcttacg ccagggtgtaa 1440
aacgtggcat cggaatggc atgcttgcgg cagagttcac gggcgggtag cccagcttcg 1500
gcttcgcgga gaatactgat gatctgttcg tcggaaaaac gcttcttcat ggggatgtcc 1560
tcatgtggct tatgaagaca ttactaacat cggggtgtac taatcaacgg ggagcaggtc 1620
accatcacta tcaatctggc aaaaattgat gcgaaaactc tcggcctgga cggttttaat 1680
atcgatggcg cgcagaaagc aaccggcagt gacctgattt ctaaatttaa agcgacagg 1740
actgataatt atcaaattaa cggtagctat aactatactg ttaattgtaga tagtggagta 1800
gtacaggata aagatggcaa acaagtttat gtgagtgtcg cggatgggtc acttacgacc 1860
agcagtgata ctcaattcaa gattgatgca actaagcttg cagtggctgc taaagattta 1920
gctcaaggta ataagattgt ctacgaaggc atcgaattta caaataccgg cactggcgct 1980
atacctgcca caggtaatgg taaattaacc gccaatgttg atggttaaggc tgttgaaattc 2040
actatttcgg ggagtgtgta tacatcaggc actagtgcga ccgttgcccc tacgacagcc 2100
ctatacaaaa atagtgcagg gcaattgact gcaacaaaag ttgaaaataa agcagcgaca 2160
ctatctgata ttgatctgaa cgctgccaag aaaacaggaa gcacgttagt tgttaacgg 2220
gcaacttacg atgttagtgc agatggtaaa acgataacgg agactgtctc tggtaacaat 2280
aaagtcatgt atctgagcaa atcagaaggc ggtagccgga ttctggtaaa cgaagatgca 2340
gcaaaatcgt tgcaatctac caccaaccgc ctcgaaacta tcgacaaagc attggctaaa 2400
gttgacaatc tgcgttctga cctcgggtgc gtacaaaacc gtttcgactc tgccatcacc 2460
aaccttggca acaccgtaaa caacctgtct tctgcccgtg gccgtatcga agatgctgac 2520

tacgcgaccg aagtgtctaa catgtctcgt gcgcagatcc tgcaacaagc gggtagctct 2580
gttctggcac aggctaacc 2599

<210> 53

<211> 1245

<212> DNA

<213> Escherichia coli

<400> 53

aacaaaaacc agtctgcgct gtcgacttct atcgagcgcc tctcttctgg tctgcgcatt 60
aacagcgcta aagatgacgc tgcgggccag gcgattgcta accgcttcac ttctaacatc 120
aaaggctctga ctcaggccgc acgtaacgcc aacgacggta tctctctggc gcagaccact 180
gaaggcgcac tgtctgaaat caacaacaac ttgcagcgtg ttcgtgaact gaccgttcag 240
gccactaccg gtactaactc tgattctgac ctgtcttcaa tccaggacga aatcaaatcc 300
cgtctcgatg aaattgaccg cgtatccggt cagactcagt tcaacggcgt gaacgtactg 360
gcaaaagatg gctcgatgaa aattcagggtc ggtgcaaagt atggtcagac aatcagcatt 420
gatttgcaga agattgattc ttctacttta gggttaaatg gtttttctgt ttcaaaaaat 480
gcagtatctg ttggtgatgc tattactcaa ttgcctggcg agacggcagc cgatgcacca 540
gtaaccatca agtttgatga ttcagtaaaa actgatttaa aactgaccga tgcttcaggg 600
ttaagtctgc ataacctcaa agatgaaaat ggtaatttaa ctaaccagta tggtgtacag 660
aatggcgga aatcttacgc tgctacagtc gctgccaatg gtaatgttac gctgaacaaa 720
gcaaatgtaa cctacagcga tgtcgcaaac ggtattgata ccgcaacgca gtcaggccag 780
ttagttcagg ttggtgcaga ttctaccggt acgcaaaaag cattcgtgtc tgtccaagg 840
aaaagctttg gcattgatga cgccgccttg aagaataaca ctggtgatgc taccgctact 900
ccaccgggaa catctgggac aacagttgtc gcagcgtaaa ttcactctgag tacgggcaaa 960
aactctgtag acgctgatgt aacggcttcc actgaattca caggtgcttc aaccaacgat 1020
ccactgactc tgctggacaa agctatcgca tctgttgata aattccgttc ttctttgggg 1080
gcggtacaga accgtctgag ctccgctgta accaacctga acaacaccac caccaacctg 1140
tctgaagcgc agtcccgtat tcaggacgcc gactatgcga ccgaagtgtc caacatgtcg 1200
aaagcgcaga ttatccagca ggcaggtaac tccgtgctgt ccaaa 1245

<210> 54

<211> 1212

<212> DNA

<213> Escherichia coli

<400> 54

aacaaaaacc agtctgcgct gtcgacttct atcgaacgcc tctcttctgg cctgcgtatt 60
aacagtgcga aagatgacgc tgccggctcag gcgatagcta accgtttcac ctctaacatt 120
aaaggcctga ctcaggctgc gcgtaacgcc aacgacggta tttctctggc gcagaccaca 180
gaaggctcgt tgtctgaaat caacaacaac ttgcaacgtg tgcgtgagtt gaccgttcag 240
gcgacgaccg gtactaactc tgattctgac ctgtcatcta ttcaggacga aatcaaatcc 300
cgtctggatg agattgaccg tgtttccggt cagaccagcgt tcaacggcgt gaatgtactg 360
gcaaaagacg gttcgatgaa gattcagggt ggcgcgaaat atggccagac tattagcatt 420
gatttacaga aaattgactc ttctacatta gggttgaatg gtttctccgt ttctgctcaa 480
tcaactaacg ttggtgattc aattactcaa attacaggag ccgctgggac aaaacctgtt 540
gggtgtgatt tcaactgctgt tgcgaaagat ctgactactg cgacaggtaa aactgtcgat 600
gtttccagcc tgacgttaca caacaccctg gatgcgaaag gggctgccac cgcacagttc 660

gtcgttcaat ccggtagtgga tttctactcc gcgtccattg accatgcaag tgggtgaagtg 720
acgttgaata aagccgatgt cgaatacaaaa gacaccgata atggactaac gactgcagct 780
actcagaaaag atcagctgat taaagttgcc gctgactctg acggcgcggc tgcgggatat 840
gtaacattcc agggtaaaaa ctacgtaca acggctccag cggcgcttaa tgatgacact 900
acggcaacag ccacagcgaa caaagttgtt gttgaattat ctacagcaac tccgactgcg 960
cagttctcag gggcttcttc tgctgatcca ctggcacttt tagacaaagc cattgcacag 1020
gttgatactt tccgctcctc cctcggtgcc gttcaaaacc gtctggactc tgcggtaacc 1080
aacctgaaca acaccaccac caacctgtct gaagcgcagt cccgtattca ggacgccgac 1140
tatgcgaccg aagtgtctaa catgtcgaaa gcgcagatca tccagcaggc gggtaactct 1200
gtgctgtcta aa 1212

<210> 55

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 55

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg agctgacggg tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttggtgag aatgacgggt aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaatgg taaagggtact 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
acgacagggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcgagg tagattatac ttacaacgct 720
aaatctggtg attttactac caccaaactc actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaatctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cgggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct taaagccgcg 1020
agcgaaggta gtgacgggtc ctctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtgggtatta ctaacgttgc cgactatata 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtagct ctgtaaatgt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtc ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatt 1680
atccagcagg ccggtaaactc cgtgctggca aaagccaacc aggtaccgca gcaggttctg 1740
tctctgctgc aggggttaa 1758

<210> 56

<211> 14024

<212> DNA

<213> Escherichia coli

<400> 56

```
gtaaccaagg gcggtacgtg cataaatttt aatgcttatac aaaactatta gcattaaaaa 60
tatataagaa attctcaaat gaacaaagaa accgtttcaa taattatgcc cgttttacaat 120
ggggccaaaa ctataatctc atcagtagaa tcaattatac atcaatctta tcaagatttt 180
gttttgtata tcattgacga ttgtagcacc gatgatacat tttcattaat caacagtcga 240
tacaaaaaca atcagaaaaat aagaatattg cgtaacaaga caaatttagg tgttgcagaa 300
agtcgaaaatt atggaataga aatggccacg gggaaatata tttctttttg tgatgcggat 360
gatttgtggc acgagaaaaa attagagcgt caaatcgaag tgttaaataa tgaatgtgta 420
gatgtggtat gttctaatta ttatgttata gataacaata gaaatattgt tggcgaagtt 480
aatgctcctc atgtgataaa ttatagaaaa atgctcatga aaaactacat aggggaatttg 540
acaggaatct ataatgccaa caaattgggt aagttttatc aaaaaaagat tggtcacgag 600
gattatttga tgtggctgga aataattaat aaaacaaatg gtgctatttg tattcaagat 660
aatctggcgt attacatgcg ttcaaataat tcactatcgg gtaataaaat taaagctgca 720
aaatggacat ggagtatata tagagaacat ttacatttgt cctttccaaa aacattatat 780
tattttttat tatatgcttc aaatggagtc atgaaaaaaa taacacattc actattaagg 840
agaaaggaga ctaaaaagtg aagtcagcgg ctaagttgat ttttttatc ctatttacac 900
tttatagtct ccagttgtat ggggttatca tagatgatcg tataacaaat tttgatacaa 960
aggatttaac tagtattata attatatctc agattttttt tgttttatta ttttatctaa 1020
cgattataaa tgaaagaaaa cagcagaaaa aatttatcgt gaactgggag ctaaagttaa 1080
tactcgtttt cctttttgtg actatagaaa ttgctgctgt agttttattt cttaaagaag 1140
gtattcctat atttgatgat gatccagggg gggctaaact tagaatagct gaaggtaatg 1200
gactttacat tagatatatt aagtattttg gtaatatagt tgtgtttgca ttaattattc 1260
tttatgatga gcataaattc aaacagagga ccatcatatt tgtatatttt acaacgattg 1320
ctttatttgg ttatcgttct gaattggtgt tgctcattct tcaatatata ttgattacca 1380
atatcctgtc aaaggataac cgtaatccta aaataaaaaag aataataggg tattttttat 1440
tggtaggggt tgtatgctcg ttgttttatc taagtttagg acaagacgga gaacaaaatg 1500
actcatataa taatatgtta aggataatta atagggttaac aatagagcaa gttgaagggtg 1560
ttccatatgt tgtttctgaa tctattaaga acgatttctt tccgacacca gagttagaaa 1620
aggaattaaa agcaataata aatagaatac aggggaataaa gcatcaagac ttattttatg 1680
gagaacgggt acataaacia gtatttggag acatgggagc aaatttttta tcagttacta 1740
cgtatggagc agaactgtta gttttttttg gttttctctg tgtattcatt atccctttag 1800
ggatatatat acctttttat cttttaaaga gaatgaaaaa aacctatagc tcgataaatt 1860
gcgcattcta ttcatatata attatgattt tattgcaata cttagtggct ggggaatgcat 1920
cggccttctt ttttggtcct tttctctcgg tattgataat gtgtactcct ctgatcttat 1980
tgcattgatac gttaaagaga ttatcacgaa atgaaaatat cagttataac tgtgacttat 2040
aataatgctg aagggttaga aaaaacttta agtagtttat caatttttaa aataaaacct 2100
tttgagatta ttatagttga tggcggctct acagatggaa cgaatcgtgt cattagtaga 2160
tttactagta tgaatattac acatgtttat gaaaaagatg aagggatata tgatgcgatg 2220
aataagggcc gaatgttggc caaaggcgac ttaatacatt atttaaaccg cggcgatagc 2280
gtaattggag atatatataa aaatatcaaa gagccatggt tgattaaagt tggccttttc 2340
gaaaatgata aacttctggg attttcttct ataaccatt caaatacagg gtattgtcat 2400
caaggggtga ttttcccaa gaatcattca gaatatgatc taaggataa aatatgtgct 2460
gattataagc ttattcaaga ggtgtttcct gaagggttaa gatctctatc tttgattact 2520
```


tcggggttatg taaaatatga tatggggggga gtatcttcaa aaaaaagaat tttaagagat 2580
 aaagagcttg ccaaaattat gtttgaaaaa aataaaaaaa accttattaa gtttattcca 2640
 atttcaataa tcaaaatttt attccctgaa cgtttaagaa gagtattgag gaaaatgcaa 2700
 tatatttgct taacttttatt cttcatgaag aatagttcac catatgataa tgaataaaat 2760
 caaaaaataa cttaaatttt gcacttttaa aaaatatgat acatcaagtg ctttaggtag 2820
 agaacaggaa aggtacagga ttatatcctt gtctgttatt tcaagtttga ttagtaaaat 2880
 actctcacta ctttctctta tattaactgt aagtttaact ttaccttatt taggacaaga 2940
 gagatttggt gtatggatga ctattaccag tcttggtgct gctctgacat ttttggaact 3000
 aggtatagga aatgcattaa caaacaggat cgcacattca tttgcgtgtg gcaaaaattt 3060
 aaagatgagt cggcaaatta gtgggtgggt cactttgctg gctggattat cgtttgtcat 3120
 aactgcaata tgctatatta cttctggcat gattgattgg caactagtaa taaaaggat 3180
 aaacgagaat gtgtatgcag agttacaaca ctcaattaaa gtctttgtaa tcatatttgg 3240
 acttggaatt tattcaaagt gtgtgcaaaa agtttatatg ggaatacaaa aagcctatat 3300
 aagtaatat gttaatgcca ttttatattt gttatctatt attactctag taatatcgct 3360
 gaaactacat gcgggactac cagttttaat tgtcagcact cttggtattc aatacatatc 3420
 gggaatctat ttaacaatta atcttattat aaagcgatta ataaagtta caaaagttaa 3480
 catacatgct aaaagagaag ctccatattt gatattaaac ggttttttct tttttatttt 3540
 acagttaggc actctggcaa catggagtgg tgataacttt ataatatcta taacattggg 3600
 tgttacttat gttgctgttt ttagcattac acagagatta tttcaaatat ctacgggtccc 3660
 tcttacgatt tataacatcc cgttatgggc tgcttatgca gatgctcatg cagcgaatga 3720
 tactcaattt ataaaaaaga cgctcagaac atcattgaaa atagtgggta tttcatcatt 3780
 cttattggcc ttcatattag tagtgttcgg tagtgaagtc gttaatattt ggacagaagg 3840
 aaagattcag gtacctgcaa cattcataat agcttatgct ttatgggtctg ttattgatgc 3900
 tttttcgaat acatttgcaa gcttttttaa tggtttgaac atagttaaac aacaaatgct 3960
 tgctgttgta acattgatat tgatcgcaat tccagcaaaa tacatcatag ttagccattt 4020
 tgggttaact gttatgttgt actgcttcat ttttatatat attgtaaatt actttatatg 4080
 gtataaatgt agttttaaaa aacatatcga tagacagtta aatataagag gatgaaaatg 4140
 aaatatatac cagttttacca accgtcattg acaggaaaag aaaaaagaata tgtaaataaa 4200
 tgtctggact caacgtggat ttcatcaaaa ggaaactata ttcagaagtt tgaaaataaa 4260
 tttgcggaac aaaaccatgt gcaatatgca actactgtaa gtaatggaac ggttgctctt 4320
 catttagctt tgttagcgtt aggtatatcg gaaggagatg aagttattgt tccaacactg 4380
 acatatatag catcagttaa tgctataaaa tacacaggag ccaccccat tttcgttgat 4440
 tcagataatg aaacttgga aatgtctgtt agtgacatag aacaaaaaat cactaataaa 4500
 actaaagcta ttatgtgtgt ccattttatac ggacatccat gtgatatgga acaaattgta 4560
 gaactggcca aaagtagaaa tttgttttga attgaagatt gcgctgaagc ctttggttct 4620
 aaatataaag gtaaatatgt gggaacattt ggagatattt ctacttttag ctttttttga 4680
 aataaaacta ttactacagg tgaaggtgga atggttgtca cgaatgacaa aacactttat 4740
 gaccgttggt tacattttta aggccaagga ttagctgtac ataggcaata ttggcatgac 4800
 gttataggct acaattatag gatgacaaat atctgcgctg ctataggatt agcccagtta 4860
 gaacaagctg atgattttat atcacgaaaa cgtgaaattg ctgatattta taaaaaaat 4920
 atcaacagtc ttgtacaagt ccacaaggaa agtaaagatg tttttcacac ttattggatg 4980
 gtctcaattc taactaggac cgcagaggaa agagaggaat taaggaaatca ccttgcatg 5040
 aaactcatcg aaacaaggcc agttttttac cctgtccaca cgatgccaat gtactcgga 5100
 aaatatcaaa agcaccctat agctgaggat cttggttggc gtggaattaa tttacctagt 5160
 ttcccagcc tatcgaatga gcaagttatt tatatttgtg aatctattaa cgaattttat 5220
 agtgataaat agcctaaaaat attgtaaagg tcattcatga aaattgcgtt gaattcagat 5280
 ggatttttac agtggggcgg tggaattgat tttattaaat atattctgtc aatattagaa 5340
 acgaaaccag aaatatgtat cgatattctt ttaccgagaa atgatataca ttctcttata 5400

agagaaaaag catttccttt taaaagtata taaaagcaa ttttaaagag ggaaaggcct 5460
 cgatggattt cattaaatag atttaatgag caatactata gagatgcctt tacacaaaat 5520
 aatatagaga cgaatcttac ctttattaaa agtaagagct ctgcctttta ttcataatttt 5580
 gatagtagcg attgtgatgt tattcttctt tgcattgcgtg ttccttcggg aaatttgaat 5640
 aaaaaagcat ggattgggta ttttatgac tttcaacact gttactatcc ttcatttttt 5700
 agtaagcgag aaatagatca aaggaatgtg ttttttaaat tgatgctcaa ttgcgctaac 5760
 aatattattg ttaatgcaca ttcagttatt accgatgcaa ataaatatgt tgggaattat 5820
 tctgcaaaac tacattctct tccatttagt ccatgccctc aattaaaatg gttcgtgat 5880
 tactctggta atattgccaa atataatatt gacaaggatt attttataat ttgcaatcaa 5940
 ttttggaac ataaagatca tgcaactgct tttagggcat ttaaaattta tactgaatat 6000
 aatcctgatg tttatttagt atgcacggga gctactcaag attatcgatt ccttgatat 6060
 tttaatgaat tgatggtttt ggcaaaaaag ctcggaattg aatcgaaaat taagatatta 6120
 gggcatatac cttaaacttga acaaattgaa ttaatcaaaa attgcattgc tgtaatacaa 6180
 ccaaccttat ttgaaggcgg gcctggaggg ggggtaacat ttgacgctat tgcattaggg 6240
 aaaaaagtta tactatctga catagatgtc aataaagaag ttaattgcgg tgatgtatat 6300
 ttctttcagg caaaaaacca ttattcatta aatgacgcga tggtaaaagc tgatgaatct 6360
 aaaatttttt atgaacctac aactctgata gaattgggtc tcaaaagacg caatgcgtgt 6420
 gcagattttt ttttagatgt tgtgaaacaa gaaattgaat cccgatctta atatattcaa 6480
 gaggtatata atgactaaag tcgctcttat tacagggtga actggacaag atggatctta 6540
 tctagctgag tttttgcttg ataaagggtg tgaagtccat ggtatcaaac gccgagcctc 6600
 atcttttaat acagaacgca tagaccatat ttatcaagat ccacatgggt ctaacccaaa 6660
 ttttcaactg cactatggag atctgactga ttcacttaac ctactagaa ttctaaagga 6720
 ggtacagcca gatgaagtat ataatttagc tgctatgagt cacgtagcag tttcttttga 6780
 gtctccagaa tatacagccg atgtcgatgc aattggtaca ttacgtttac tggagcaat 6840
 tcgcttttta ggattggaaa acaaaacgcg tttctatcaa gcttcaacct cagaattata 6900
 tggacttggt caggaaatcc ctcaaaaaga atccaccctc ttttatcctc gttcccctta 6960
 tgcagttgca aaactttacg catattggat cagggtaaat tatcgagagt catatgggat 7020
 ttatgcatgt aatgggtatat tgttcaatca tgaatctcca cgccgtggag aaacgtttgt 7080
 aacaaggaaa attactcgag gacttgcaaa tattgcacaa ggcttggaat catgtttgta 7140
 tttaggggat atggattcgt tacgagattg gggacatgca aaagattatg ttagaatgca 7200
 atgggtgatg ttacaacagg agcaaccgga agattttgtg attgcaacag gagtccaata 7260
 ctacgtccgt cagtttgtcg aaatggcagc agcacaactt ggtattaaga tgagctttgt 7320
 tggtaaagga atcgaagaaa aaggcattgt agattcggtt gaaggacagg atgctccagg 7380
 tgtgaaacca ggtgatgtca ttgttctgtg tgatcctcgt tatttccgac cagctgaagt 7440
 tgatactttg cttggagatc cgagcaaagc taatctcaaa cttgggttga gaccagaaat 7500
 tactcttgct gaaatgattt ctgaaatggt tgccaaagat cttgaagccg ctaaaaaaca 7560
 ttctctttta aaatcgcatg gtttttctgt aagcttagct ctggaatgat gatgaataag 7620
 caacgtattt ttattgctgg tcaccaagga atgggtggat cagctattac ccgacgcctc 7680
 aaacaacgtg atgatgttga gttggtttta cgtactcggg atgaattgaa cttgttggat 7740
 agtagcgctg ttttgattt tttttcttca cagaaaatcg accaggttta tttggcagca 7800
 gcaaaagtcg gaggtatttt agctaacagt tcttatcctg ccgattttat atatgagaat 7860
 ataatgatag aggcgaatgt cattcatgct gccacaaaaa ataattgtaa taaactgctt 7920
 ttctcgggtt cgtcgtgtat ttatcctaag ttagcacacc aaccgattat ggaagacgaa 7980
 ttattacaag ggaaacttga gccacaacaa gaaccttatg ctatcgcaaa aattgcagg 8040
 attaaattat gtgaatctta taaccgtcag tttgggcgtg attaccgttc agtaatgcca 8100
 accaatcttt atgggtccaaa tgacaatttt catccaagta attctcatgt gattccggcg 8160
 cttttgcgcc gctttcatga tgctgtggaa aacaattctc cgaatgttgt tgtttgggga 8220
 agtggctact caaagcgtga attcttacat gtagatgata tggcttctgc aagcatttat 8280

gtcattggaga tgccatacga tatatggcaa aaaaatacta aagtaatgtt gtctcatatc 8340
 aatattggaa caggtattga ctgcacgatt tgtgagcttg cggaaacaat agcaaaagtt 8400
 gtaggttata aagggcatat tacgttcgat acaacaaagc ccgatggagc ccctcgaaaa 8460
 ctacttgatg taacgcttct tcatcaacta ggttggaatc ataaaattac ccttcacaag 8520
 ggtcttgaaa atacatacaa ctgggttctt gaaaaccaac ttcaatatcg ggggtaataa 8580
 tgtttttaca ttcccaagac tttgccacaa ttgtaaggct tactcctctt atttctatag 8640
 atttgattgt ggaaaacgag tttggcgaaa ttttgctagg aaaacgaatc aaccgcccgg 8700
 cacagggcta ttggttcggt cctgggtgta ggggtgtgaa agatgaaaaa ttgcagacag 8760
 cctttgaacg attgacagaa attgaactag gaattcgttt gcctctctct gtgggtaagt 8820
 tttatggat ctggcagcac ttctacgaag acaatagtat ggggggagac ttttcaacgc 8880
 attatatagt tatagcattc cttcttaaat tacaaccaa cattttgaaa ttaccgaagt 8940
 cacaacataa tgcttattgc tggctatcgc gagcaaagct gataaatgat gacgatgtgc 9000
 attataattg tcgcgcata ttttaacaata aaacaaatga tgcgattggc ttagataata 9060
 aggatataat atgtctgat cgccaataat tgctgtagtt atggccggtg gtacaggcag 9120
 tcgtctttgg ccactttctc gtgaactata tccaaagcag tttttacaac tctctggtga 9180
 taacaccttg ttacaaacga ctttgctacg actttcaggg ctatcatgtc aaaaaccatt 9240
 agtgataaca aatgaacagc atcgctttgt tgtggctgaa cagttaagg aaataaataa 9300
 attaaatggt aatattattc tagaaccatg cgggcgaaat actgcaccag caatagcgat 9360
 atctgcgttt catgcgttaa aacgtaatcc tcaggaagat ccattgcttc tagttcttgc 9420
 ggcagaccac gttatagcta aagaaagtgt tttctgtgat gctattaaaa atgcaactcc 9480
 catcgcta at caaggtaaaa ttgtaacgtt tggaattata ccagaatatg ctgaaactgg 9540
 ttatgggtat attgagagag gtgaactatc tgtaccgctt caagggcatg aaaatactgg 9600
 tttttattat gtaaataagt ttgtcgaaaa gcctaactgt gaaaccgcag aattgtatat 9660
 gacttctggt aatcactatt ggaatagtgg aatattcatg ttttaaggcat ctgtttatct 9720
 tgaggaattg agaaaattta gacctgacat ttacaatgtt tgtgaacagg ttgcctcatc 9780
 ctcatacatt gatctagatt ttattcgatt atcaaaagaa caatttcaag attgtcctgc 9840
 tgaatctatt gattttgctg taatggaaaa aacagaaaaa tgtgttgtat gcctgttgga 9900
 tattgggttg agtgacgttg gatcttggca atcgttatgg gacattagtc taaaatcgaa 9960
 aacaggagat gtatgtaaag gtgatattt aacctatgat actaagaata attatatcta 10020
 ctctgagtca gcgttggtag ccgccattgg aattgaagat atggttatcg tgcaactaa 10080
 agatgccgtt cttgtgtcta aaaagagtga tgtacagcat gtaaaaaaaaa tagtcgaaat 10140
 gcttaaatg cagcaacgta cagagtatat tagtcatcgt gaagttttcc gaccatgggg 10200
 aaaatttgat tcgattgacc aaggtgagcg atacaaagtc aagaaaatta ttgtgaaacc 10260
 tggtaggggg ctttctttaa ggatgcatca ccacgttctt gaacattgga tcgtgcttct 10320
 tggtagacga aaagtaaccc ttggcgataa aactaaacta gtcaccgcaa atgaatcgat 10380
 atacattccc cttggcgag cgtatagtct tgagaatccg ggcataatcc ctcttaattct 10440
 tattgaagtc agttcagggg attatttggg agaggatgat attataagac agaaagaacg 10500
 ttacaaacat gaagattaac atatgaaatc tttaacctgc tttaaagcct atgatattcg 10560
 cgggaaatta ggcgaagaac tgaatgaaga tattgcctgg cgcatggggc gtgcctatgg 10620
 cgaatttctc aaaccgaaaa ccattgtttt aggcggtgat gtcgcctca ccagcgaagc 10680
 gttaaaactg gcgcttgca aaggtttaca ggatgcgggc gtcgatgtgc tggatatcgg 10740
 tatgtccggc accgaagaga tctatttcgc caggttccat ctccgagtg atggcggcat 10800
 cgaagttacc gccagccata accgatgga ttacaacggc atgaagctgg tgcgcgaagg 10860
 ggctcgccc atcagcggtg ataccggact gcgcgatgtc cagcgtctgg cagaagccaa 10920
 tgacttccct cctgtcgatg aaaccaaagc tggctgctat cagcaaatca atctgcgtga 10980
 cgttacgtt gatcacctgt tcggttatat caacgtcaaa aacctcacgc cgctcaagct 11040
 ggtgatcaac tccgggaacg gcgcagcggg tccggtgggt gacgccattg aagcccgatt 11100
 taaagccctc ggcgaccgg tggaaatta caaagtacac aacacgccgg acggcaattt 11160

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| ccccaacggt | attcctaacc | cgctgctgcc | ggaatgccgc | gacgacaccc | gtaatgcggt | 11220 |
| catcaaacac | ggcgcgata | tgggcattgc | ctttgatggc | gattttgacc | gctgtttcct | 11280 |
| gtttgacgaa | aaagggcagt | ttatcgaggg | ctactacatt | gtcggcctgc | tggcagaagc | 11340 |
| gttccctcgaa | aaaaatcccg | gcgcgaagat | catccacgat | ccacgtctct | cctggaacac | 11400 |
| cgttgatgtg | gtgactgccg | caggcgccac | cccggtaatg | tcgaaaaccg | gacacgcctt | 11460 |
| tattaaagaa | cgtatgcgca | aggaagacgc | catctacggt | ggcgaaatga | gcgctcacca | 11520 |
| ttacttccgt | gatttcgctt | actgcgacag | cggcatgata | ccgtggctgc | tggcgccga | 11580 |
| actggtgtgc | ctgaaaggaa | aaacgctggg | cgaaatgggtg | cgcgaccgga | tggcggcgtt | 11640 |
| tccggcaagc | ggtgagatca | acagcaaact | ggcgcaaccc | gttgaggcaa | ttaatcgcg | 11700 |
| ggaacagcat | tttagccgcg | aggcgctggc | ggtggatcgc | accgatggca | tcagcatgac | 11760 |
| ctttgccgac | tggcgcttta | acctgcgctc | ctccaacacc | gaaccgggtg | tgcggttgaa | 11820 |
| tgtggaatca | cgcgggtgatg | taaagcta | ggaaaagaaa | actaaagctc | ttcttaaatt | 11880 |
| gctaagttag | tgattattta | cattaatcat | taagcgtatt | taagattata | ttaaagta | 11940 |
| gttattgcgg | tatatgatga | atatgtgggc | ttttttatgt | ataacgacta | taccgcaact | 12000 |
| ttatctagga | aaagattaat | agaaataaag | ttttgtactg | accaatttgc | atttcacgtc | 12060 |
| acgattgaga | cgttcccttg | cttaagacat | tttttcatcg | cttatgta | aacaaatgtg | 12120 |
| ccttatataa | aaaggagaac | aaaatggaac | ttaaataat | tgagacaata | gatttttatt | 12180 |
| atccctgttt | acgatattat | agccaaagtt | gtatcctgca | tcagtctctgc | aatatttcac | 12240 |
| gagtgccttg | ttaactgaat | acatgtctgc | cattttccag | atgataacga | cgctcatcgca | 12300 |
| attgatggta | aaacacttcg | gcacacttat | gacaagagtc | gtcgcagagg | agtggttcat | 12360 |
| gtcattagt | cgtttcagca | atgcacagtc | tggctcctcg | atagatcaag | acggatgaga | 12420 |
| aacctaattgc | gttcacagtt | attcatgaac | tttctaaaat | gatgggtatt | aaaggaaaaa | 12480 |
| taatcataac | tgatgcgatg | gcttgccaga | aagatatgtc | agagaagata | taaaaacaga | 12540 |
| gatgtgatta | tttattcgct | gtaaaaggaa | ataagagtcg | gcttaataga | gtctttgagg | 12600 |
| agatatttac | gctgaaagaa | ttaaataatc | caaaacatga | cagttacgca | attagtga | 12660 |
| agaggcacgg | cagagacgat | gtccgtcttc | atattgtttg | agatgctcct | gatgagctta | 12720 |
| ttgatttcac | gtttgaatgg | aaagggctgc | agaatttatg | aatggcagtc | cactttctct | 12780 |
| caataatagc | agagcaaaaag | aaagaatccg | aatgacgat | caaataattat | attagatctg | 12840 |
| ctgctttaac | cgcagagaag | ttcgccacag | taaatcgaaa | tcactggcgc | atggagaata | 12900 |
| agttgcacag | tagcctgatg | tggtaatgaa | tgaaatcgac | tataatataa | gaaggcgagt | 12960 |
| tgcattcgaa | tgattttcta | gaatgcggca | catcgctatt | aatatctgac | aatgataatg | 13020 |
| tattcaaggc | aggattatca | tgtaagatgc | gaaaagcagt | catggacaga | aacttcctag | 13080 |
| cgtcaggcat | tgcagcgtgc | gggctttcat | aatcttgc | tggttttgat | aagatatttc | 13140 |
| tttgagatg | ggaaaatgaa | tttgtatggt | atttttggtg | ctggaagtta | tggtagagaa | 13200 |
| acaataccca | ttctaaatca | acaaataaag | caagaatgtg | gttctgacta | tgctctggtt | 13260 |
| tttgtggatg | atgttttggc | aggaaagaaa | gttaatgggt | ttgaagtgc | ttcaaccaac | 13320 |
| tgctttctaa | aagcccctta | tttaaaaaag | tattttaatg | ttgctattgc | taatgataag | 13380 |
| atacgacaga | gagtgtctga | gtcaatatta | ttacacgggg | ttgaaccaat | aactataaaa | 13440 |
| catccaaata | gcgttggtta | tgatcatact | atgataggta | gtggcgctat | tatttctccc | 13500 |
| tttgttacaa | tatctactaa | tactcatata | gggagggttt | ttcatgcaa | catatactca | 13560 |
| tacgttgac | atgattgtca | aataggagac | tatgttacat | ttgctcctgg | ggctaaatgt | 13620 |
| aatggatag | ttgttattga | agacaatgca | tatataggct | cgggtgcagt | aattaagcag | 13680 |
| ggtgttccta | atcgccact | tattattggc | gcgggagcca | ttataggtat | gggggctggt | 13740 |
| gtcactaaaa | gtgttcctgc | cggtataact | gtgtgcggaa | atccagcaag | agaaatgaaa | 13800 |
| agatcgccaa | catctattta | atgggaatgc | gaaaacacgt | tccaaatggg | actaatgttt | 13860 |
| aaaatatata | taatttcgct | aatttactaa | attatggctt | ctttttaagc | tatcctttac | 13920 |
| ttagttatta | ctgatacagc | atgaaattta | taatactctg | atacattttt | atacgttatt | 13980 |
| caagccgcat | atctagcgg | aaccctgac | aggagtaa | aatg | | 14024 |

- 47 -

<210> 57

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 57

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccggt ttactttctaa cattaaaggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtattttctg ttgctgcagac caccgaaggc 240
gctgtgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggccact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
gatgaaattg accgctgata cgccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacggttcaa tgaaaattca ggttggtgctg aatgacggcg aaaccatcac gatcgacctg 480
aaaaaaaaatc attctgatac tctgggtctg aatggcttta acgtaaatgg taaagggtact 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcaca gttggcgggc tagattatac ttacaacgct 720
aaatctggtg attttactac cactaaatct actgctggta cgggtgtaga cgccgcggcg 780
caggctgctg attcagcttc aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaatctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cgggtggaagc caggcatatc tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct caaagcagcg 1020
agcgaaggta gtgacggtgc ctctctgaca ttcaatggca cagaatatac catcgcaaaa 1080
gcaactcctg cgacaaccac tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtggtatca ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaatgt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg caatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc cgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1680
attcagcagg ccggttaact cgtgctggca aaagctaacc aggtaccgca gcaggttctg 1740
tctctgctgc aggggttaa

```

1758

<210> 58

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 58

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgcagcggg tcaggcgatt gctaaccggt ttactttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtattttctg ttgctgcagac caccgaaggc 240

```

```

gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggccact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacggttcaa tgaaaattca ggttggtgcg aatgacggcg aaaccatcac gatcgacctg 480
aaaaaaatcg attctgatac tctgggtctg aatggcttta acgtaaatgg taaagggtact 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcaca gttggcgggc tagattatac ttacaacgct 720
aaatctggtg attttactac cactaaatct actgctggta cgggtgtaaa cgccgcggcg 780
caggctgctg attcagcttc aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaactctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cggtggaagc caggcatagc tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct caaagcagcg 1020
agcgaaggta gtgacgggtg ctctctgaca ttcaatggca cagaatatac catcgcaaaa 1080
gcaactcctg cgacaaccac tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtgggtatca ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggcactg ctgtaaatgt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg caatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcggtcac caacctgaac aacaccacta ccaacctgtc cgaagcgcag 1620
tcccgatttc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1680
atccagcagg ccggtaaactc cgtgctggca aaagctaacc aggtaccgca gcagggttctg 1740
tctctgctgc agggttaa

```

1758

<210> 59

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 59

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg agctgacggt tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgctctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttggtgcg aatgacgggt aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaatgg taaagggtact 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcgggc tagattatac ttacaacgct 720
aaatctggtg attttactac caccaaactc actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaactctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900

```

```

tcagcaggta atatcaccat cgggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgctgct taaagccgcg 1020
agcgaaggta gtgacggtgc ctctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtggtatta ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaagt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgt ggttctgcaa cgaccaaccc gcttgcgtcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatt 1680
atccagcagg ccggttaactc cgtgctggca aaagccaacc aggtaccgca gcaggttctg 1740
tctctgctgc aggggttaa

```

1758

<210> 60

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 60

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gtaaccgtt ttacttctaa cattaaggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggccact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacgggtcaa tgaaaattca ggttggtgcg aatgacggcg aaaccatcac gatcgacctg 480
aaaaaaaaatc attctgatac tctgggtctg aatggcttta acgtaaattg taaaggctact 540
attaccaaca aagctgcaac ggtaagtgt ttaacttctg ctggcgcgaa gttaaacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcaca gttggcggcg tagattatac ttacaacgct 720
aaatctggtg attttactac cactaaatct actgctggta cgggtgtaga cgccgcggcg 780
caggctgctg attcagcttc aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaatctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cgggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgctgct caaagcagcg 1020
agcgaaggta gtgacggtgc ctctctgaca ttcaatggca cagaatatac catcgcaaaa 1080
gcaactcctg cgacaaccac tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtggtatca ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaagt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgt ggttctgcaa cgaccaaccc gcttgcgtcc 1500
ctggacgacg caatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560

```

cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc cgaagcgcag 1620
 tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1680
 attcagcagg ccggttaactc cgtgctggca aaagctaacc aggtaccgca gcaggttctg 1740
 tctctgctgc aggggttaa 1758

<210> 61

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 61

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
 aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
 gcgaaggatg acgccgcagg tcaggcgatt gctaaccgtt ttacttctaa cattaaggc 180
 ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcgcagac caccgaaggc 240
 gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggccact 300
 acaggggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
 gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
 gacggttcaa tgaaaattca ggttggtgcg aatgacggcg aaaccatcac gatcgacctg 480
 aaaaaaatcg attctgatac tctgggtctg aatggcttta acgtaaattg taaagggtact 540
 attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
 acgacagggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
 gataaattag ggaatggcga taaagtcaca gttggcggcg tagattatac ttacaacgct 720
 aaatctggtg attttactac cactaaatct actgctggta cgggtgtaga cgccgcggcg 780
 caggctgctg attcagcttc aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
 ggtaaactctg ttaatgggtc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
 tcagcaggta atatcaccat cggtggaagc caggcatacg tagacgatgc aggcaacttg 960
 acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgctgct caaagcagcg 1020
 agcgaaggta gtgacgggtg ctctctgaca ttcaatggca cagaatatac catcgcaaaa 1080
 gcaactcctg cgacaaccac tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
 gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
 attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
 gatgctgcga agtcttatgt ggatgataaa ggtggtatca ctaacgttgc cgactataca 1320
 gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
 actgatacca ataaagatta tgctccagca attggcactg ctgtaaattg gaactccgcg 1440
 ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
 ctggacgacg caatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
 cgtctggatt ccgcggtcac caacctgaac aacaccacta ccaacctgtc cgaagcgcag 1620
 tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1680
 atccagcagg ccggttaactc cgtgctggca aaagctaacc aggtaccgca gcaggttctg 1740
 tctctgctgc aggggttaa 1758

<210> 62

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 62

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccggt ttactttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtattttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg agctgacggt tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttggtgag aatgacgggt aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaattg taaaggtagt 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaacacc 600
acgacagggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcgagg tagattatac ttacaacgct 720
aaatctgggtg attttactac caccaaactc actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaatctg ttaatggttc ttacaccaca aaagatggta ctggtttctt cgaaacggat 900
tcagcaggta atatcaccat cggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct taaagccgcg 1020
agcgaaggta gtgacgggtc ctctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat ttcttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtggtatta ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaattg gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtc ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgag 1620
tcccgattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgagatt 1680
atccagcagg ccggtaaactc cgtgctggca aaagccaacc aggtaccgca gcaggttctg 1740
tctctgctgc aggggttaa 1758

<210> 63

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 63

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcagg tcaggcgatt gctaaccggt ttactttctaa cattaaggc 180
ctgactcagg cggcccgtaa cgccaacgac ggtattttctg ttgcgcagac caccgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtattcgtg aactgacggt tcaggctcact 300
acagggacta actccgattc tgacctggac tccatccagg acgaaatcaa atctcgtctt 360
gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt gctggcgaaa 420
gacggttcaa tgaaaattca ggttggtgag aatgacggcg aaaccatcac gatcgacctg 480
aaaaaaatcg attctgatac tctgggtctg aatgggttta acgtaaattg taaaggtagt 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaacacc 600
acgacagggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660

gataaattag ggaatggcga taaagtcaca gttggcggcg tagattatac ttacaacgct 720
aaatctggtg attttactac cactaaatct actgctggta cgggtgtaga cgccgcggcg 780
caggctgctg attcagcttc aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaactctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cggtggaagc caggcatagc tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct caaagcagcg 1020
agcgaaggta gtgacggtgc ctctctgaca ttcaatggca cagaatatac catcgcaaaa 1080
gcaactcctg cgacaaccac tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgca agtcttatgt ggatgataaa ggtggtatca ctaacgttgc cgactataca 1320
gtctcttaca gcgttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaagt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg caatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt ccgcagtcac caacctgaac aacaccacta ccaacctgtc cgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatc 1680
attcagcagg ccggtaaact cgtgctggca aaagctaacc aggtaccgca gcaggttctg 1740
tctctgctgc agggttaa 1758

<210> 64

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 64

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gtaaccgtt ttacttctaa cattaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac caccgaaggc 240
gcgctgtctg aaatcaacaa caacttacag cgtatccgtg agctgacggt tcaggcttct 300
accggaacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctt 360
gatgaaattg accgcgtatc cggccagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttgggtgcg aatgacggtg aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaattg taaaggtact 540
attaccaaca aagctgcaac ggtaagtgt ttaacttctg ctggcgcgaa gttaaacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcggcg tagattatac ttacaacgct 720
aaatctggtg attttactac caccaaactc actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaactctg ttaatggttc ttacaccaca aaagatggta ctgtttcttt cgaaacggat 900
tcagcaggta atatcaccat cggtggaagc caggcatagc tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct taaagccgcg 1020
agcgaaggta gtgacggtgc ttctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat tacttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgca agtcttatgt ggatgataaa ggtggtatta ctaacgttgc cgactataca 1320

```

gtctctttaca gcggttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaagt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt cgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatt 1680
atccagcagg ccggttaactc cgtgctggca aaagccaacc aggtaccgca gcagggttctg 1740
tctctgctgc agggttaa
1758

```

<210> 65

<211> 1758

<212> DNA

<213> Escherichia coli

<400> 65

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttactttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg agctgacggg tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggttgggtgc aatgacgggt aaactatcac tatcgacctg 480
aagaaaatcg attctgatac tctgggtctg aatgggttta acgtaaattg taaagggtact 540
attaccaaca aagctgcaac ggtaagtgat ttaacttctg ctggcgcgaa gttaaaccacc 600
acgacaggtc tttatgatct gaaaaccgaa aataccttgt taactaccga tgctgcattc 660
gataaattag ggaatggcga taaagtcacc gttggcggcg tagattatac ttacaacgct 720
aaatctgggtg attttactac caccaaatct actgctggta cgggtgtaga cgccgcggcg 780
caggctactg attcagctaa aaaacgtgat gcgttagctg ccacccttca tgctgatgtg 840
ggtaaattctg ttaattggttc ttacaccaca aaagatggta ctggttcttt cgaaacggat 900
tcagcaggta atatcaccat cggtggaagc caggcatacg tagacgatgc aggcaacttg 960
acgactaaca acgctggtag cgcagctaaa gctgatatga aagcgtgct taaagccgcg 1020
agcgaaggta gtgacgggtg ctctctgaca ttcaatggca ctgaatatac tatcgcaaaa 1080
gcaactcctg cgacaacctc tccagtagct ccgttaatcc ctggtgggat ttcttatcag 1140
gctacagtga gtaaagatgt agtattgagc gaaaccaaag cggctgccgc gacatcttca 1200
attaccttta attccggtgt actgagcaaa actattgggt ttaccgcggg tgaatccagt 1260
gatgctgcga agtcttatgt ggatgataaa ggtgggtatta ctaacgttgc cgactataca 1320
gtctctttaca gcggttaacaa ggataacggc tctgtgactg ttgccgggta tgcttcagcg 1380
actgatacca ataaagatta tgctccagca attggtactg ctgtaaagt gaactccgcg 1440
ggtaaaatca ctactgagac taccagtgtt ggttctgcaa cgaccaaccc gcttgctgcc 1500
ctggacgacg ctatcagctc catcgacaaa ttccgttctt ccctgggtgc tatccagaac 1560
cgtctggatt cgcagtcac caacctgaac aacaccacta ccaacctgtc tgaagcgcag 1620
tcccgtattc aggacgccga ctatgcgacc gaagtgtcca acatgtcgaa agcgcagatt 1680
atccagcagg ccggttaactc cgtgctggca aaagccaacc aggtaccgca gcagggttctg 1740
tctctgctgc agggttaa
1758

```

<210> 66

<211> 1788

<212> DNA

<213> Escherichia coli

<400> 66

```

atggcacaag tcattaatac caacagcctc tcgctgatca ctcaaaataa tatcaacaag 60
aaccagtctg cgctgtcgag ttctatcgag cgtctgtctt ctggcttgcg tattaacagc 120
gcgaaggatg acgccgcggg tcaggcgatt gctaaccgtt ttacttctaa cattaaaggc 180
ctgactcagg ctgcacgtaa cgccaacgac ggtatttctg ttgcacagac cactgaaggc 240
gcgctgtccg aaatcaacaa caacttacag cgtatccgtg agctgacggg tcaggcttct 300
accgggacta actctgattc ggatctggac tccattcagg acgaaatcaa atcccgtctc 360
gacgaaattg accgcgtatc cggtcagacc cagttcaacg gcgtgaacgt actggcaaaa 420
gacggttcga tgaaaattca ggtagggtcg aacgacggcc agactatcac tattgatctg 480
aagaaaattg actctgatac gctggggctg aatgggttta acgtgaatgg ttccgggtacg 540
atagccaata aagcggcgac cattagcgac ctgacagcag cgaaaatgga tgctgcaact 600
aatactataa ctacaacaaa taatgcgctg actgcatcaa aggcccttga tcaactgaaa 660
gatggtgaca ctgttactat caaagcagat gcagctcaaa ctgccacggg ctatacatat 720
aatgcatctg ctggtaactt ctcattcagt aatgtatcga ataatacttc agcaaaagca 780
ggatgatgtg cagctagcct tctcccgcg gctgggcaaa ctgctagtgg tgtttacaaa 840
gcagcaagcg gtgaagtga ctttgatgtt gatgcgaatg gtaaaattac aatcggagga 900
caggaagcct atttaactag tgatggtaac ttaactacaa acgatgctgg tgggtgcgact 960
gcggctacgc ttgatggttt attcaagaaa gctgggtgat gtcaatcaat cgggtttaat 1020
aagactgcat cagtcacgat ggggggaaca acttataact ttaaaacggg tgctgatgct 1080
ggtgctgcaa ctgctaacgc aggggtatcg ttcactgata cagctagcaa agaaaccgtt 1140
ttaaataaag tggctacagc taaacaaggc acagcagttg cagctaacgg tgatacatcc 1200
gcaacaatta cctataaatc tggcgttcag acgtatcagg cggatatttg cgcaggtgac 1260
ggtactgcta gcgcaaaaata tgccgataat actgacgttt ctaatgcaac agcaacatac 1320
acagatgctg atggtgaaat gactacaatt ggttcataca ccacgaagta ttcaatcgat 1380
gctaacaacg gcaaggtaac tgttgattct ggaactgggt cgggtaaata tgcgccgaaa 1440
gtcggggctg aagtatatgt tagtgcta atggtacttta caacagatgc aactagcgaa 1500
ggcacagtaa caaaagatcc actgaaagct ctggatgaag ctatcagctc catcgacaaa 1560
ttccgttcat ccctgggggc tatccaaaac cgtttggtt cgcgcgtcac caacctgaac 1620
aacaccacta ccaacctgtc tgaagcgcag tcccgtattc aggacgccga ctatgcgacc 1680
gaagtgtcca acatgtcgaa agcgcagatt atccagcagg ccggttaactc cgtgctggca 1740
aaagccaacc aggtaccgca gcaggttctg tctctactgc aggggttaa 1788

```

<210> 67

<211> 1398

<212> DNA

<213> Escherichia coli

<400> 67

```

aacaaatctc agtcttctct tagctctgct attgagcgtc tgtcttctgg tctgcgtatt 60
aacagcgcaa aagacgatgc agcaggtcag gcgattgcta accgttttac ggcaaatatt 120
aaagggtctga cccaggcttc ccgtaacgca aatgatggta tttctgttgc gcagaccact 180
gaagggtgcgc tgaatgaaat taacaacaac ctgcagcgta ttcgtgaact ttctgttcag 240
gcaactaacg gtactaactc tgacagtgc ctgacctcca tccagtccga aatccagcag 300
cgtctgagtg aaattgaccg tgtttctggg cagactcagt ttaacggcgt taaagtgctg 360
gcttctgata aggatatgac tattcaggtt ggtgcaaacg acggcgaaac aattactatt 420

```

```

aaactgcagg aaattaattc cgacacactg ggattatctg gttttggtat taaagatcct 480
actaaattaa aagccgcaac ggctgaaaca acctatcttg gatcgacagt taagcttgct 540
gacgctaata cacttgatgc agatattaca gctacagtta aaggcactac gactccgggc 600
caacgtgacg gtaatattat gtctgatgct aacggtaagt tgtacgttaa agttgccggg 660
tcagataaac ccgctgaaaa tgggtattat gaagttactg tggaggatga tccgacatct 720
cctgatgcag gtaagctgaa gctgggggct ctacggggtg cccagcctca agctggtaat 780
ttaaaggaag tcacaacggg gaaaggggag ggggctattg atgttcagtt ggggtactgat 840
accgcaaccg cttctatcac aggtgcaaaa ctctttaagt tagaagacgc caatggcaaaa 900
gatactgggt catttgcggt gattgggtgat gacggtaaac agtatgcagc gaatgttgat 960
cagaaaacag gagcagtttc cgttaaaaca atgtcttaca ctgatgctga cgggtgtcaa 1020
cacgacaatg ttaaagttga actgggtgga agcgatggca aaaccgaagt tgtaactgca 1080
accgatggca aaacttacag tgtagtgat ttacaaggta agagcctgaa aactgattct 1140
attgcagcaa tttctacgca gaaaacagaa gatcctttgg ctgctatcga taaagcactg 1200
tctcaggttg actcgttgcg ttctaacctt ggtgcaattc aaaatcgttt cgactctgcc 1260
atcaccaacc ttggcaacac cgtaaacacac ctgtcttctg cccgtagccg tatcgaagat 1320
gctgactacg cgaccgaagt gtctaacatg tctcgtgcgc agatcctgca acaagcgggt 1380
acctctgttc tggcgcag

```

1398

<210> 68

<211> 1479

<212> DNA

<213> Escherichia coli

<400> 68

```

aacaaatctc agtcttctct gagctccgcc attgaacgct tctcttctgg cctgcgtatt 60
aacagtgcta aagatgacgc agcaggctcag gcgattgcta accgttttac agcaaatatt 120
aaagggtctga ctcaggcttc ccgtaacgcg aatgatggta tttctgttgc gcagaccact 180
gaagggtgcgc tttctgaaat caacaataac ttacagcgta ttcgtgaatt gtcagtacag 240
gccactaatg gtacaaactc tgactccgac ctgaattcaa ttcaggatga aattacacaa 300
cgcttagtg aaattgatcg tgtttctaac cagacacaat ttaatgggtg aaaagttctg 360
gcttctgac agactatgaa aattcaagta ggtgcgaacg atgggtgaaac cattgagatt 420
gcccttgata aaattgatgc taaaaccttg gggcttgata actttagcgt agcaccagga 480
aaagttccaa tgcctctgc ggttgcaatt aagagcgaag ccgctcctga cttaactaag 540
gtaaattgcaa ctgatggtag tgtgggaggt gctaaagcat tcggtagcaa ttataaaaaat 600
gctgatgttg aaacttattt tggtagccgt aatgtacaag atacaaagga tacaactgat 660
gcgaccggta ctgcaggaaac aaaagtttat caagtacagg tggaagggca gacttatttt 720
gttggtcaag ataataatac caacacgaac ggttttacat tattgaaaca aaactctaca 780
ggttatgaaa aagttcaggt ggggtggttaag gatgttcagt tagcaaactt tgggtggtcgt 840
gtaactgcat ttgttgaaga taatgggttct gccacatcag ttgatttagc tgcgggtaaa 900
atgggttaaag cattagctta taatgatgca ccaatgtctg tttattttgg gggaaaaaac 960
ctagatgtcc accaagtaca agatacccaa gggaaatcctg tacctaattc atttgctgct 1020
aaaacatcag acggcaccta cattgcagta aatgtagatg ccgctacagg taacacgtct 1080
gttattactg atcctaattg taaggcagtt gaatgggcag taaaaaatga tggttctgca 1140
caggcaatta tgcgtgaaga tgataagggt tatacagcca atatcacgaa taagacggca 1200
accaaagggt ctgaactcag tgcctcagat ttgaaagcct tagcaaccac aaatccatta 1260
tccacattag acgaagcttt ggcaaaaagt gataagttgc gcagttcttt ggggtgcagta 1320
caaaaccgtt tcgactctgc catcaccaac cttggcaaca ccgtaaacaa cctgtcttct 1380
gcccgtagcc gtatagaaga tgctgactac gcaaccgaag tgtctaacat gtctcgtgcg 1440

```

WO 99/61458

- 56 -

PCT/AU99/00385

cagatcctgc aacaagcggg tacctctggt ctggcacag

1479

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU 99/00385**A. CLASSIFICATION OF SUBJECT MATTER**Int Cl⁶: C07H 21/04, (C12Q 1/10, 1/68, C12R 1:19), G01N 37/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN: Chemical Abstracts, Medline and Derwent World Patents Index using keywords e. coli and flagellin
ANGIS: Sequences corresponding to Fig 25, Fig 10 and Fig 34 (random sample)**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| P,X | JOURNAL OF BACTERIOLOGY, (1999 Jan) 181 (1), Reid S D; Selander R K; Whittam T S, "Sequence diversity of flagellin (fliC) alleles in pathogenic Escherichia coli", 153-160 | 1-33 |
| X | MOLECULAR MICROBIOLOGY (1994) 12(2), Tominaga A; Mahmoud T; Enomoto M, "Molecular characterization of intact, but cryptic, flagellin genes in the genus Shigella", 277-285 | 1-33 |
| X | JOURNAL OF MOLECULAR BIOLOGY (1994) 238, Fahrner K A; Block S M; Krishnaswamy S; Parkinson J S; Berg H C, "A mutant hook-associated protein (HAP3) facilitates torsionally induced transformations of the flagellar filament", 173-186 | 1-33 |

☒ Further documents are listed in the continuation of Box C☐ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
27 July 1999Date of mailing of the international search report
- 3 AUG 1999Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200
WODEN ACT 2606
AUSTRALIA
Facsimile No.: (02) 6285 3929

Authorized officer

MATTHEW FRANCIS
Telephone No.: (02) 6283 2424

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU 99/00385

| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|-----------------------|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | SCIENCE (Washington DC) (1997) 277(5331), Blattner F R et al, "The complete genome sequence of Escherichia coli K-12", 1453-1462 | 1-33 |
| X | JOURNAL OF BACTERIOLOGY (Sept. 1993) 175(17), Schoenhals G; Whitfield C, "Comparative analysis of flagellin sequences from Escherichia coli strains possessing serologically distinct flagellar filaments with a shared complex surface pattern", 5395-5402 | 1-33 |
| X | JOURNAL OF BACTERIOLOGY (Feb. 1998) 180(4), Ratiner Y A, "New flagellin-specifying genes in some Escherichia coli strains", 979-984 | 1-33 |
| X | FEMS MICROBIOLOGY LETTERS (1987) 48, Ratiner Y A, "Different alleles of the flagellin gene hagB in Escherichia coli standard H test strains", 97-104 | 1-33 |
| X | FEMS MICROBIOLOGY LETTERS (1985) 29, Ratiner Y A, "Two genetic arrangements determining flagellar antigen specificities in two diphasic Escherichia coli strains", 317-323 | 1-33 |